Transmission Optimization Schemes for Multicarrier CDMA Systems

HU Fan

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Information Engineering

Acknowledgement

First of all, I would like to express my deep gratitude to my supervisor, Professor Lok Tat-Ming for introducing this research topic to me, giving me guidance and encouragement during the whole research period and helping me to polish this thesis. In helping me as a beginner in doing research in wireless communication field, Professor Lok patiently explained some basic concepts to me and teach me very useful research methods. I would like to thank him for his support and trust during these two years.

With this opportunity, I would like to thank Professor Yue On-Ching and Professor Ng Wai-Yin for some very helpful discussions. Many thanks go to my friends who give me help and support. These include Mr. Chen Chung-Shue, Mr. Kwan Ho-Yuet, Mr. Huang Yi, Mr. Zhao Fei, Mr. Zheng Ming, Ms. Li Feng-Jun, Mr. Ng Cho-Yiu and Mr. Tse Hok-Man.

Last but not the least, I would like to thank my family for their endless love and support, especially to my parents.

Abstract

The new millennium is starting with tremendous anticipation of the potential that wireless and mobile technologies will bring to us. The successful deployment of wireless voice communication systems and the merging of wireless communications and computer networks promise a bright future for wireless high data rate services such as Internet access or multimedia applications. One of the major problems in high data rate transmission is multipath fading. Multicarrier (MC) CDMA has been proposed as an effective and robust technique to combat multipath fading. In the past few years, different MC-CDMA systems have been proposed and investigated. MC-CDMA system with direct sequence spreading is considered to derive the benefits of both multicarrier modulation and direct sequence spreading.

In MC-CDMA systems, each user is assigned one or more codes for transmission. Since the fading channel is time varying and frequency selective, the subcarriers that appear to be in deep fade for one user may not be in deep fade for others. Therefore, a user may experience different performance with different codes and different time slots. In the previous work, the optimal code assignment scheme is proposed for MC-CDMA system. It has been shown that significant performance gain can be achieved by applying optimal code assignment. In order to further utilize the time sharing characteristic of wireless channel, we investigate the optimal assignment of codes with multiple time slots to minimize the transmission time. For practical

implementation, we also propose a suboptimal algorithm. Since each channel can be time shared by users, we introduce the concept of channel sharing factor. From the simulation results, we can see that with suitable value of channel sharing factor, the system performance will be near optimal.

Considering longer period of time, we need to extend the previous assignment scheme to certain scheduling policy. Since wireless channel is time varying, we can schedule users to transmit at those codes and times that are relatively good to them. Joint scheduling and resource allocation over multiple wireless channels is proposed. However, the formulated problem is not a convex optimization. Thus we use an suboptimal algorithm with low computational complexity, which still can achieve very good system performance.

摘 要

在新千年，人們對無線通信技術給日常生活所能帶來的變化充滿了期望。無線語音系統的巨大成功以及無線網絡與傳統計算機網絡的成功融合為無線高速率業務，如 Internet 接入、多媒體應用等，提供了光明的前景。 然而，高速率數據傳輸要克服的一個主要問題是多徑衰落。多載波 CDMA 是對抗多徑衰落的一種有效、穩健的技術。在過去的幾年中，人們提出並深入研究了多種不同的多載波 CDMA 技術。其中，與直接序列擴頻相結合的多載波 CDMA 技術被認為同時繼承了多載波調製和直接序列擴頻的雙重優點。

在多載波 CDMA 系統中，每個用戶會被分配一個或多個碼。衰落信道是時變的且頻率選擇性的，對某個用戶呈現深度衰落的子載波對於其他用戶卻未必依然是深度衰落。因此，用戶對於不同的碼及不同的時隙將表現出完全不同的性能。在過去的研究工作中，人們已提出多載波 CDMA 系統的最優碼分配。最優碼分配使得系統性能獲得巨大的提高。為了進一步利用無線信道的時間共享特性，本文研究了碼及時隙的最優分配。基於實用的考慮，本文還提出了一個次最優算法。因為信道在時間域上可以被用戶共享，所以我們引入了信道共享係數的概念。從仿真結果來看，通過適當選擇信道共享係數，系統性能可以接近最優。

考慮一個比較長的時間段，我們將先前的分配方案擴展為一種調度算法。因為無線信道是時變的，所以系統可以調度用戶在合適的碼及時隙上傳輸數據。本文提出了一種聯合調度及資源分配方案。然而，建模後的問題不是凸規劃問題。因而，我們使用一種低複雜度的次最優算法，而且仍然可以獲得非常好的系統性能。

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Evolution of Mobile Communications

At the end of the 19th century, electromagnetic waves were initially discovered as a wireless communication medium. Mobile car phone service was firstly introduced in the late 1940s. The early mobile systems were inhibited by their low capacity and limited mobility. In the late 1970s and early 1980s, analog cellular systems were developed. They are much improved in the system capacity and user mobility. These cellular systems provide voice transmissions only. The Advanced Mobile Phone System (AMPS), Nordic Mobile Telephone (NMT), and Total Access Communication System (TACS) are the well-known 1G systems, which are based on frequency division multiplexing (FDM) techniques [1].

Second generation (2G) wireless systems have employed digital modulation, advanced semiconductor technology and improved call processing techniques to have better transmission quality, system capacity, and coverage. Compared to 1G systems, spectrum-efficient schemes such as TDMA or CDMA techniques are used in 2G systems. The major standards include Global System for Mobile Communications (GSM, Europe, previously named as Groupe Spécial Mobile), Personal Digital Cellular (PDC, Japan), IS-136 (US, TDMA) and IS-95 (US, CDMA) [2]. Although 2G systems mainly provide wireless voice service, data transmissions are supported too. There are various 1G and 2G standards used in different countries or regions. Most of them are incompatible.

In 1992, globalization of cellular mobile standards is suggested by ITU (International Telecommunications Union). Worldwide radio spectrum of frequencies around 2 GHz is allocated for 3G radio systems. The standardization on IMT-2000 (International Mobile Telecommunications 2000) framework has been carried out to provide seamless wideband (bit rates up to 2 Mbps) multimedia services. Different radio transmission technologies (RTT) have been proposed for IMT-2000. Harmonization of these air interface proposals results in the recommendations of CDMA-based wideband cellular 3G solutions, WCDMA and cdma2000, as the major future wireless mobile communications standards [1]. CDMA has been accepted as the major 3G multiple access scheme worldwide. The development of 3G standards is now going on in the two 3G partnership projects joined by multiple regional standards developing organizations (SDO). 3GPP works on WCDMA standardization. This is specified as UMTS (Universal Mobile Telecommunication System) in 3GPP. 3GPP2 works on cdma2000 standard. As a result of convergence, the global wideband CDMA standard has adopted three operation modes including Multi-Carrier (MC) CDMA, Direct Spread (DS) CDMA and Time Division Duplex (TDD) CDMA. MC-CDMA is based on cdma2000 while the DS-CDMA and TDD-CDMA modes are based on UTRA FDD (UMTS Terrestrial Radio Access, Frequency Division Duplex) and UTRA TDD specifications respectively. UTRA FDD proposal is mainly based on wideband DS-CDMA scheme while UTRA TDD is mainly based on wideband TD-CDMA scheme [3]. In this thesis, we will have discussions on the bandwidth allocation in multi-rate TD-CDMA and wideband CDMA based on the UTRA model.

1.2 Overview of Multicarrier Systems

With the prosperity of wireless systems, wireless multimedia applications become a part of our daily life. However current wireless systems cannot meet the consumer's growing expectation on various service and high data rate. So in the next-generation wireless systems, it is expected that different classes of traffics are supported and the data rate are highly improved to provide the multimedia service, which are drastically different from the existing second-generation wireless systems. The proposed third generation wireless standards UMTS/IMT-2000 uses wide-band CDMA (WCDMA) to address the higher rate requirements of multimedia applications [3]. One of the major problems in high data rate transmission is multipath fading of the wireless channel. Multicarrier (MC) systems are proven to be more immune to transmission impairment in wireless channel than single carrier systems [4], [5]. Given the physical nature of the wireless fading channel, frequency selective fading is commonly encountered. Multicarrier modulation (MCM) is demonstrated to be an effective way to combat the negative effects of fading by dividing the frequency selective fading channel into a number of flat fading sub-channels corresponding to the carrier frequencies. Let us divide the channel with a bandwidth of B Hz into M small sub-carriers, spaced by B/M Hz. The spectrum of the different sub-carriers mutually overlaps and the signals on different sub-carriers are orthogonal, giving therefore an optimal efficiency with small adjacent channel interference. More and more applications, such as broadcast of digital audio, digital television and wireless local area networks (LAN), are proposed with MCM [4].

Another advantage of MCM is the possibility of efficient fast Fourier Transform (FFT) implementations, which is called Orthogonal Frequency Division Multiplexing (OFDM). OFDM is a matured multicarrier modulation scheme, which transmits parallel data streams over orthogonal sub-carriers. In 1971, Weinstein and Ebert simplified the implementation of OFDM systems with Discrete Fourier Transform (DFT) methods [6]. Since then, OFDM is adopted in many applications such as the IEEE 802.11a and the European HIPERLAN/2 standard for high-speed WLANs.

Combining OFDM and CDMA techniques, a novel class of multicarrier communication schemes called multicarrier CDMA (MC-CDMA) emerged in 1993. Much attention has been drawn on this modulation scheme due to its numerous advantages, such as enhancement of robustness against frequency selective fading and high scalability in possible data transmission rate. MC-CDMA system with direct sequence spreading is considered to derive the benefits of both multicarrier modulation and direct sequence spreading [7].

1.3 Outline of This Thesis

In this thesis, we will investigate the resource allocation in multicarrier CDMA systems. Chapter 2 gives a survey of the multicarrier communications. After a brief introduction of the concept of multicarrier modulation and its advantages against single carrier modulation, we review the basics of OFDM and the main elements in OFDM transceivers. We also introduce the DFT implement of OFDM systems. Then, we review the three types of multicarrier CDMA communication systems, which emerged in 1993 as a combination of OFDM and CDMA. We discuss the advantages and disadvantages of these three multicarrier CDMA schemes. In Chapter 3, optimization problems with different objective, such as transmission time, throughput and total transmission power, are considered. Fortunately, all the optimization problems are convex programming or can be converted into convex programming. We give our solutions for the two cases of short and long period of time respectively. We introduce queueing model into our problem in Chapter 4. After some transformation and approximation, the original problem can be simplified into an optimization problem, but still nonconvex. For practical concern, a suboptimal algorithm, which is the combination of the power water-filling and code reassignment, is proposed. Simulation results show that significant performance gain can be achieved over other scheduling policies.

Chapter 2

Multicarrier Communication Systems

2.1 Introduction

The principle of multicarrier communications, which is to transmit data by dividing it into several parallel and interleaved bit streams, and using these to modulate several carriers, has its origin at least thirty yeas ago, when it was applied in the Collins Kineplex system [4]. Recently, much attention has been drawn to high-rate high-quality wireless mobile communications. In hostile wireless channels, such as those in urban area, because of the frequency selective fading, broadband transmission with single carrier modulation has severe performance degradation due to inter symbol interference (ISI) effect. Multicarrier modulation is therefore considered as an alternative to fulfill such a stringent service requirement [8]. In multicarrier systems, the broadband channel is divided into N narrowband sub-channels that are ISI-free as long as N is sufficiently large. Data steam is also divided into parallel sub-steams. Then, it will be modulated and sent out over these narrow band sub-channels. Because the bandwidth of the sub-channel is usually smaller than the fading coherent bandwidth, the narrow band signals do not experience frequency selective fading. With this advantage, unlike the single carrier modulation with very complex adaptive equalizer, simple receiver with high performance can be adopted in multicarrier communication systems, which make it very suitable for high-rate wireless mobile communications.

2.2 Multicarrier Modulation (MCM) Scheme versus Single Carrier Modulation (SCM) Scheme

The MCM scheme is designed to mitigate ISI by dividing a wideband channel into a number of narrow band sub-channels that are less susceptible to ISI. The rate of the incoming signal R is reduced by a factor of N by increasing the symbol period by N times. In the frequency domain, the bandwidth of the R/N bit steams is N times smaller than the initial bandwidth of the signal. Figure 2.1 shows the R bps data rate signal with period T and baseband bandwidth B, while Figure 2.2, shows one of the R/N bps bit steams with period NT and baseband bandwidth B/N [8].

Figure 2.1: Time and frequency domain representation of information signal

Figure 2.2: Time and frequency domain representation of sub-channel signal

Hence, after passing a serial to parallel converter, the original data stream is divided into N sub-streams and each bit sub-stream is then modulated to one sub-carrier. The modulated signals are summed together and transmitted over the channel. Figure 2.3 shows a typical multicarrier modulation transmitter with a Quadrature Amplitude Modulation (QAM) modulator.

As shown by Figure 2.3, with baseband bandwidth B/N, each sub-channel has a maximum bandwidth of twice its baseband bandwidth, $2B/N$, after pulse shaping and modulation to the passband. It should be noted that only the in-phase components of the signal are shown in Figure 2.3. The corresponding quadrature component of the signal is simultaneously modulated by sine wave carriers.

Figure 2.4 compares a single carrier (SCM) and a multicarrier modulation (MCM) scheme [8]. Here, B_{SCM} and B_{MCM} denote the bandwidths of transmitted SCM and MCM signals, respectively. For MCM, f_k, $F_k(f;t)$, N and Δf denote the frequency of the kth subcarrier, the frequency spectrum of pulse waveform of the kth subcarrier, the total number of the subcarri-

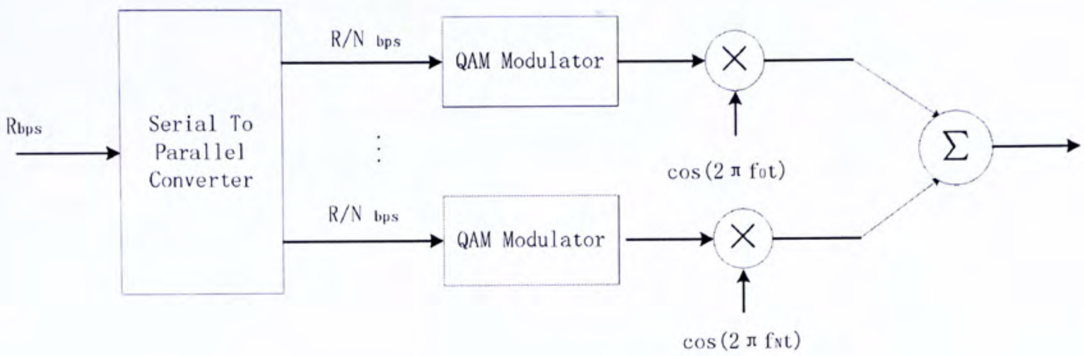

Figure 2.3: Transmitter design of a multicarrier modulation system

ers, and subcarrier separation, respectively. The frequency spectrum of the MCM signal can be written as [8]

$$U_{MCM}(f;t) = \sum_{k=1}^{N} F_k(f;t) \tag{2.1}$$

Through a frequency selective fading channel with transfer function $H(f;t)$, the frequency spectra of received SCM and MCM signals are of this form [8]

$$R_{SCM}(f;t) = H(f;t)U_{SCM}(f;t) \tag{2.2}$$

$$\begin{aligned} R_{MCM}(f;t) &= H(f;t)U_{MCM}(f;t) \\ &= \textstyle\sum_{k=1}^{N} H_k(f;t)F_k(f;t) \end{aligned} \tag{2.3}$$

where $U_{SCM}(f;t)$ is the frequency spectrum of transmitted SCM signal and $H_k(f;t)$ is the channel transfer function corresponding to the narrow band B_k, which is the frequency band occupied by the kth subcarrier. When the number of subcarrier is sufficiently large, the amplitude and phase response of $H_k(f;t)$ can be assumed as constant over B_k. As shown in Figure 2.4, we can use a line, which will becomes flat as N increases, to approximate the

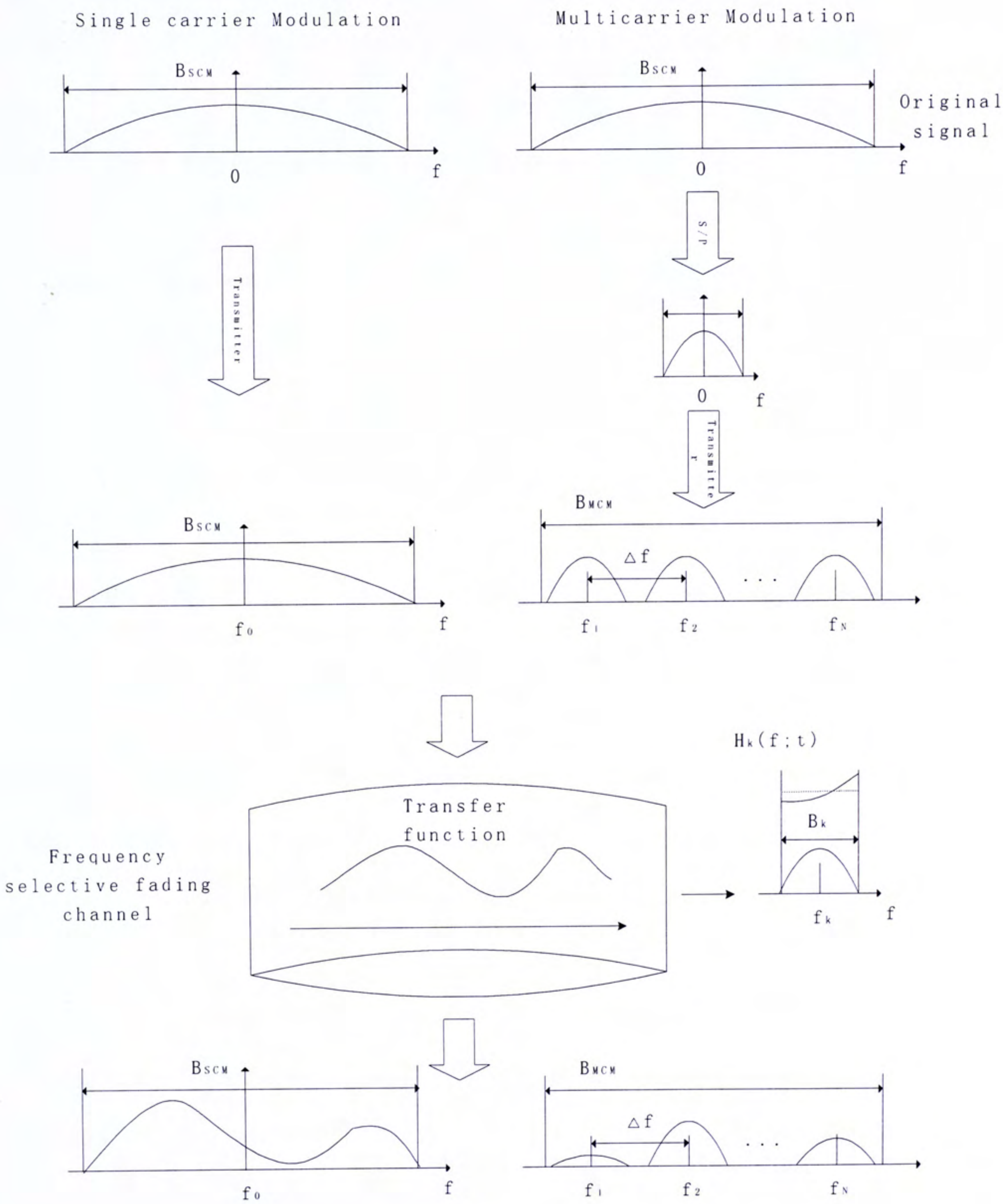

Figure 2.4: Comparison of SCM and MCM

curve of $H_k(f;t)$. Therefore, $R_{MCM}(f;t)$ can be approximated as [8]

$$R_{MCM}(f;t) = \sum_{k=1}^{N} H_k(t)F_k(f;t) \tag{2.4}$$

where $H_k(t)$ is the complex representation pass loss over the narrow band B_k.

Equation (2.4) and Figure 2.4 clearly show that MCM is effective and robust to combat frequency selective fading. Hence, MCM requires no equalization or at most one-tap equalization for each subcarrier, whereas SCM need complicated adaptive equalization [8].

2.3 Orthogonal Frequency Division Multiplexing (OFDM) Systems

If the sub-channel's frequency separation, Δf, is greater than the bandwidth of each subcarrier, $2B/N$, the spectrum of the multicarrier modulated signal will not have overlapped subcarriers, which is called Frequency Division Multiplexing (FDM) or Nonoverlapping Frequency Division Multiplexing (NFDM) [9]. The disadvantage of this kind of technique is that at least twice of the original bandwidth R bps is needed to transmit the multicarrier modulated signal.

However, if the carriers are separated by $1/NT$, where T is the duration of the original signal, the subcarriers overlap and the same bandwidth as the original signal is needed to transmit the modulated data. Adjacent car-

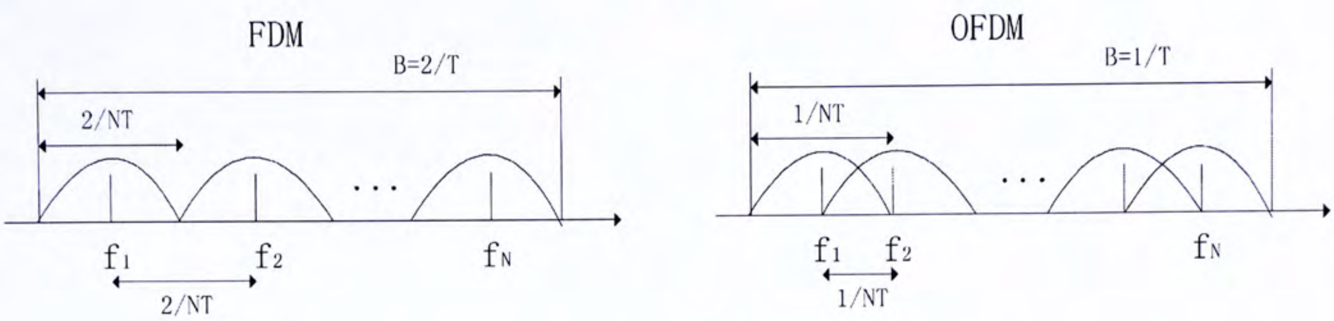

Figure 2.5: Frequency spectrum of FDM and OFDM

rier separation of $1/NT$ guarantees the orthogonality of the subcarriers so that the receiver can accurately separate the data of each subcarrier. This overlapping multicarrier modulation is called Orthogonal Frequency Division Multiplexing (OFDM) [9]. We use Figure 2.5 to show the frequency spectrum of FDM and OFDM in order to further make clear these concepts.

In 1971, Weinstein and Ebert showed that a digital implementation of an OFDM system can be based on Discrete Fourier Transform (DFT) methods [6], which eliminate arrays of sinusoidal generator and coherent demodulation required in parallel data systems and make the implementation of the technology cost effective. However, it still took more than a decade from that time until OFDM start to attract broad interest, boosted by the demand for broadband communication systems and the vast development in digital signal processing technology. Before we investigate the utilization of OFDM in today's communication systems, let's review how DFT can be used to implement OFDM [9].

A block diagram of the OFDM transmitter implemented by DFT is given in Figure 2.6. Consider a data sequence $[d_0, d_1, ..., d_{N-1}]$, where each d_i is a

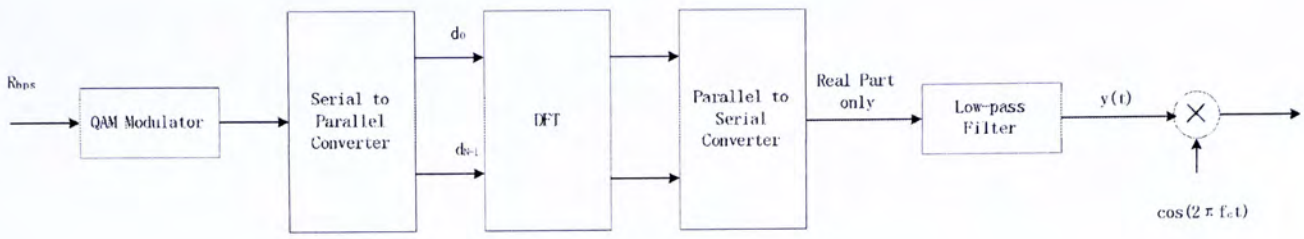

Figure 2.6: Block diagram of OFDM transmitter with DFT technology

complex number $d_n = a + jb_n$. If we perform a DFT on the vector $\{d_n\}_{n=0}^{N-1}$, the result is a vector $S = \{S_0, S_1, \cdots, S_{N-1}\}$ of N complex numbers with the form as follows [9]

$$\begin{aligned} S_m &= \sum_{n=0}^{N-1} d_n e^{-j(2\pi nm/N)} \\ &= \sum_{n=0}^{N-1} d_n e^{-j2\pi f_n t_m}, \quad m = 0, 1, \cdots, N-1 \end{aligned} \tag{2.5}$$

where $f_n \triangleq \frac{n}{N\Delta t}$ and $t_m \triangleq m\Delta t$. Δt is an arbitrarily chosen time interval. The real part of the vector S has components [9]

$$Y_m = \sum_{n=0}^{N-1} a_n \cos 2\pi f_n t_m + b_n \sin 2\pi f_n t_m, \quad m = 0, 1, \cdots, N-1 \tag{2.6}$$

If these components are applied to a low-pass filter at time interval $m\Delta t$, a signal is obtained that closely approximates the original OFDM signal [9]

$$y(t) = \sum_{n=0}^{N-1} a_n \cos 2\pi f_n t + b_n \sin 2\pi f_n t, \quad 0 \le t \le N\Delta t \tag{2.7}$$

Demodulation at the receiver is also carried out via the DFT of a vector of samples of the received signal. Figure 2.7 shows the structure of a conventional OFDM receiver.

Because only the real part of the original OFDM signal is transmitted, it is necessary to sample twice as fast as expected, i.e., at intervals $\Delta t/2$.

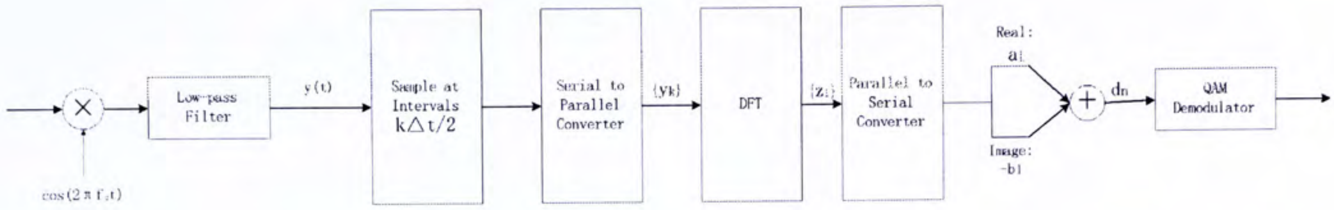

Figure 2.7: Block diagram of OFDM receiver with DFT technology

When there is no channel distortion, the total $2N$ samples of the received signal is as follows [9]

$$Y_k = y(k\frac{\Delta t}{2}) = \sum_{n=0}^{N-1} (a_n \cos\frac{2\pi nk}{2N} + b_n \sin\frac{2\pi nk}{2N}), \quad k = 0, 1, \cdots 2N-1 \tag{2.8}$$

The DFT performed on sequence $\{Y_k\}_{k=0}^{2N-1}$ yields [9]

$$z_l = \frac{1}{2N}\sum_{k=0}^{2N-1} Y_k e^{-j(2\pi lk/2N)} = \begin{cases} a_0, & l = 0 \\ \frac{a_l - jb_l}{2}, & l = 1, 2, \cdots N-1 \\ irrelevant, & l > N-1 \end{cases} \tag{2.9}$$

where the equality

$$\frac{1}{2N}\sum_{k=0}^{2N-1} e^{j(2\pi mk/2N)} = \begin{cases} 1, & m = 0, \pm 2N, \pm 4N, \cdots \\ 0, & otherwise \end{cases} \tag{2.10}$$

has been employed. As shown by equation (2.9), the original data a_l and b_l are available as the real and image part of z_l. Hence, the original d_n can be recovered. After passing the QAM demodulator, we can get the source symbols with data rate of R bps.

Now, we can see that OFDM communication systems can be readily built up with the DFT technology. Furthermore, if the number of subcarriers N is power of 2, many Fast Fourier Transform (FFT) algorithms can be adopted to perform the DFT transform, which will further reduce the system complexity [9].

Due to recent advances of digital signal Processing (DSP) and Very Large Scale Integrated circuit (VLSI) technologies, the initial obstacles of OFDM implementation such as massive complex computation, and high speed memory do not exist anymore. Today, OFDM is widely utilized both in wireless communications and for cable based data transmission. In the latter field, the technology is normally called Discrete Multitone (DMT) and applied in the new xDSL standards like Asymmetric DSL (ADSL) [8]. In wireless transmission, ODFM was adopted by the IEEE 802.11a and the European HIPERLAN/2 standard for high-speed WLANs. Moreover, OFDM was chosen for starting up terrestrial digital broadcasting services like the Digital Audio Broadcasting (DAB) systems and Digital Video Broadcasting (DVB) projects [9].

2.4 Multicarrier CDMA

Code Division Multiple Access (CDMA) is a multiplexing technique where a number of users simultaneously and asynchronously access a channel by spreading each user's information with distinct signature sequences. In 1993, an epoch of CDMA application, there types of new multiple access schemes based on a combination of CDMA and OFDM were proposed. N.Yee, J-P

Linnartz and G. Fettweis [10], K. Fazel and L. Papke [11], and A. Chouly, A. Brajal and S. Jourdan [12] proposed Multicarrier (MC-) CDMA. V. M. Dasilva and E. S. Sousa [13] proposed Multicarrier DS-CDMA. L. Vandendorpe [14] proposed Multitone (MT-) CDMA. Soon, Multicarrier CDMA becomes a very hot topic and lots of research works have done on it. In this section, we will review these three types of Multicarrier CDMA schemes and discuss their advantages and disadvantages in terms of the transmitter and receiver structures.

2.4.1 MC-CDMA

Unlike the conventional CDMA that spreads the original data stream using the pre-assigned spreading code in the time domain, the MC-CDMA transmitter spreads user's data stream over different subcarriers using a given spreading code in the frequency domain. In other word, each subcarrier is modulated by a chip of the user's spreading code. In [8], K. Fazel and G. Fettweis have shown that spreading codes like the Hadamard codes are optimum in maintaining orthogonality between subcarriers and reducing intermodulation in nonlinear amplifiers.

We consider a system of totally M users. We define the vector $\mathbf{C}_m = [C_{m,1}, C_{m,2}, \cdots, C_{m,N}]^T$ as the spreading code for the mth user. By choosing M orthogonal vectors, we can construct the MC-CDMA transmitter for totally M users. Without loss of generality, we depicted the structure of the MC-CDMA transmitter of the mth user in Figure 2.8. All users' signals are

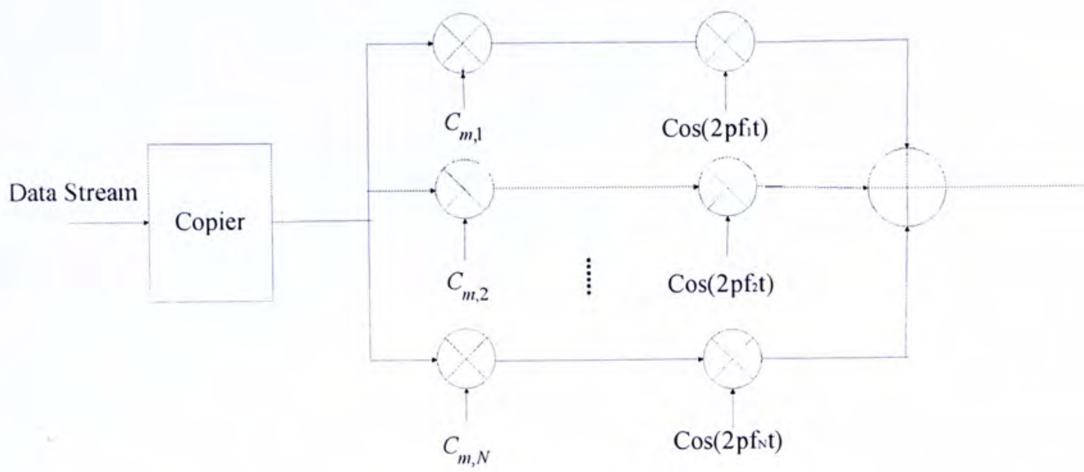

Figure 2.8: Block diagram of the MC-CDMA transmitter for one user

summed together, modulated the carrier frequency f_c, and sent out through the wireless mobile channel. If we assume the channel is distortion-free, without loss of generality, we consider using the receiver as shown by Figure 2.9 to receive and recover the data stream of the mth user. The transmitter and receiver for other users can be readily obtained by adopting their own spreading code, i.e., appropriately change the subscripts of the spreading code in Figure 2.8 and Figure 2.9 [8].

2.4.2 MC-DS-CDMA

The MC-DS-CDMA transmitter spreads the serial/parallel converted data stream using a given spreading code in the time domain so that the resulting spectrum of each subcarrier can satisfy the orthogonality condition with the minimum frequency separation [13]. We consider the mth user and assume $\mathbf{C}_m = [C_{m,1}, C_{m,2}, \cdots, C_{m,N}]^T$ as its spreading code. As shown by Figure 2.10, the symbols modulated on the N carriers are summed together before

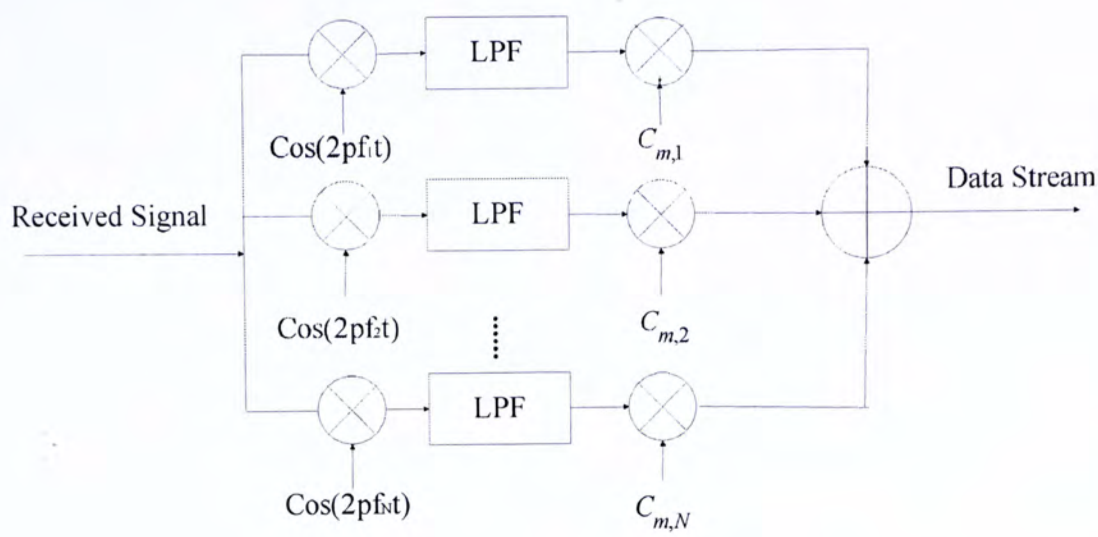

Figure 2.9: Block diagram of the MC-CDMA receiver for one user

being transmitted over the channel. We use the receiver shown by Figure 2.11 for the data reception of the mth user. If we assume the wireless channel has no distortion, all users' data will maintain their orthogonality. Therefore, at the receiver end, we use the spreading code $\mathbf{C}_m$ and low-pass filter to pick up the data stream of the mth user, while signals of other users are gotten rid of as white noise. After passing a parallel-to-serial converter, we can get the original data of the mth user.

2.4.3 MT-CDMA

The MT-CDMA scheme is similar to MC-DS-CDMA in the sense that the incoming bit stream of one user is divided into N sub-streams and every sub-stream is spread in time domain with the same signature sequence pre-assigned to that particular user. Unlike MC-DS-CDMA, MT-CDMA uses longer spreading codes in proportion to the number of subcarriers. After

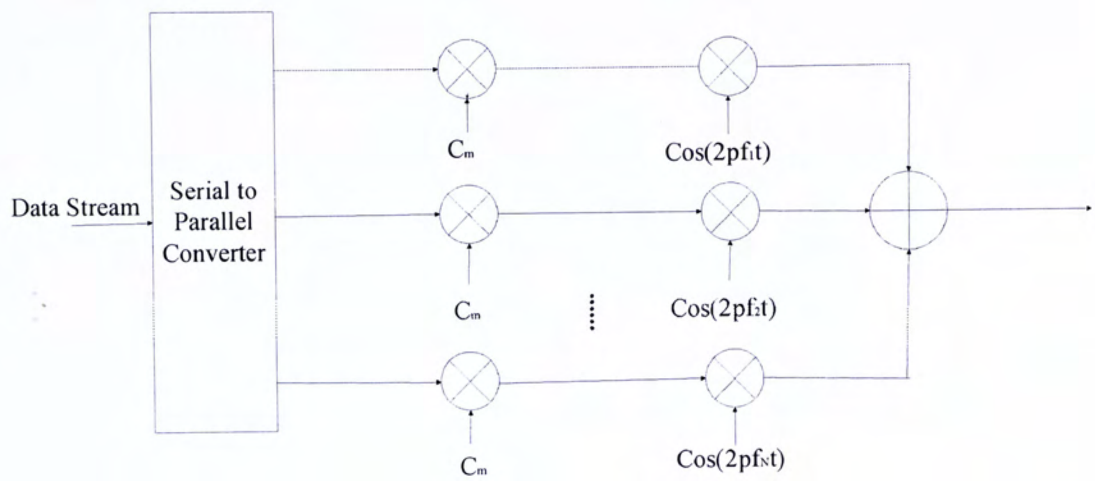

Figure 2.10: Block diagram of the MC-DS-CDMA transmitter for one user

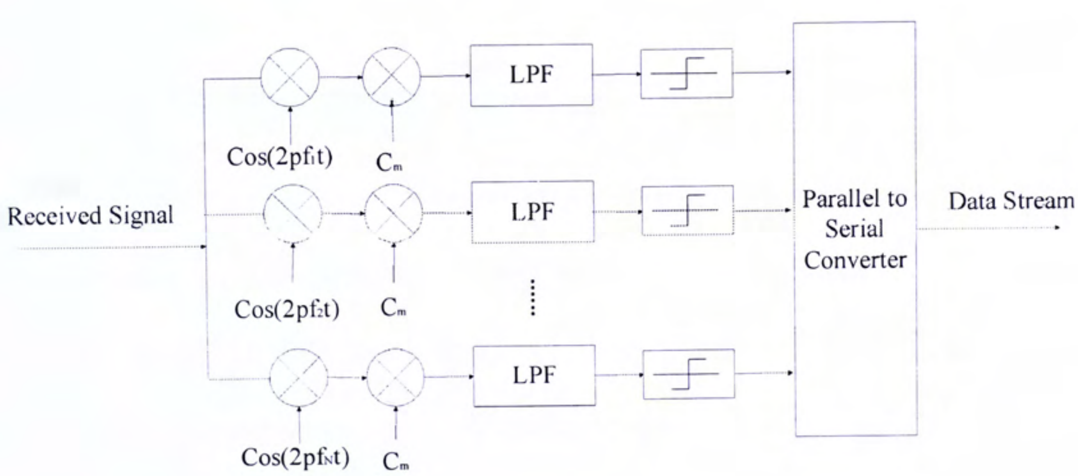

Figure 2.11: Block diagram of the MC-DS-CDMA receiver for one user

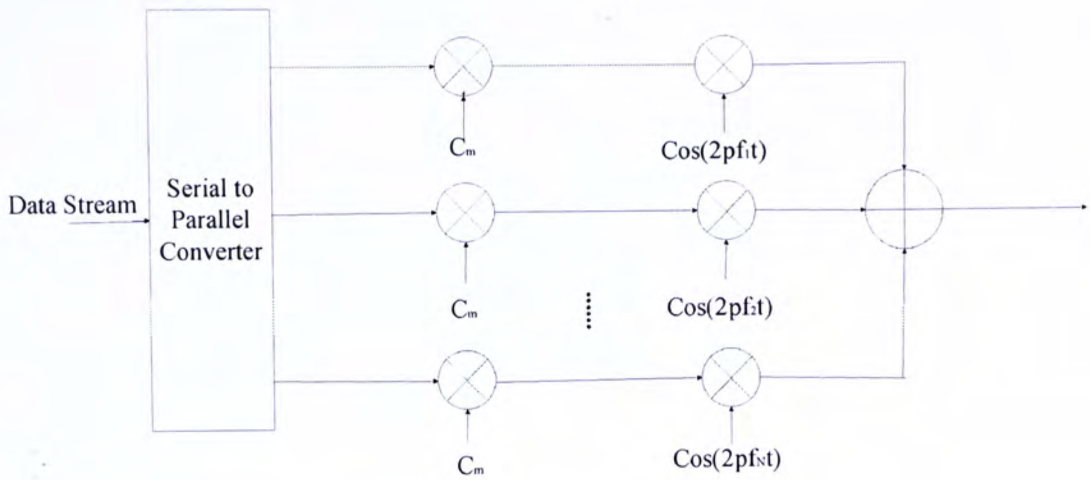

Figure 2.12: Block diagram of the MT-CDMA transmitter for one user

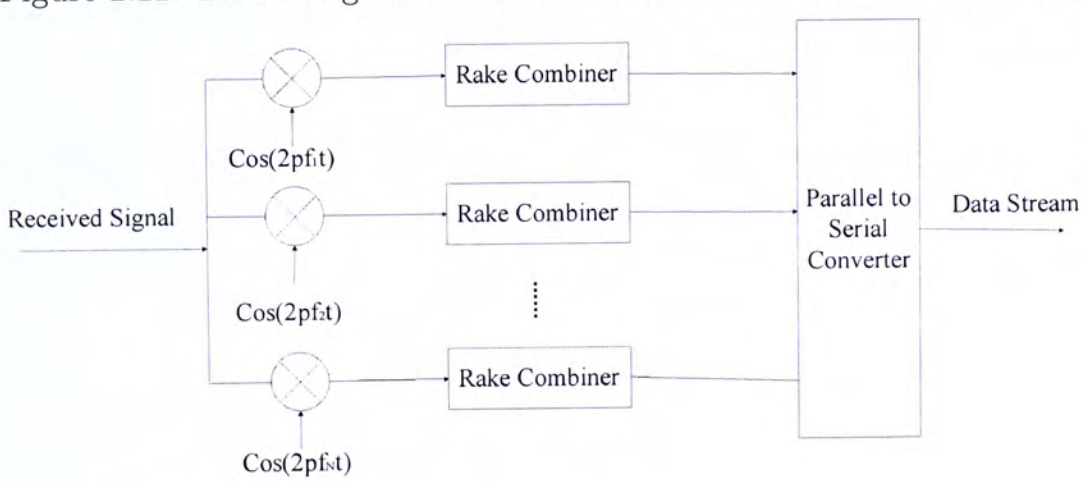

Figure 2.13: Block diagram of the MT-CDMA receiver for one user

spreading, the data on each subcarrier are no longer orthogonal. Therefore, when we consider the design of the receiver, Rake Combiner with the same structure as DS-CDMA Rake Receiver must be used to receive the data stream of each subcarrier [14]. Without loss of generality, Figure 2.12 and Figure 2.13 show the structure of MT-CDMA transmitter and receiver of the mth user, respectively. Other users' transmitter and receiver can be readily designed by appropriately changing the subscripts.

Chapter 3

Optimization for MC-CDMA Systems

3.1 System Model

In MC-CDMA systems [15], each user is assigned one or more codes for transmission. Since the fading channel is time varying and frequency selective, the subcarriers that appear to be in deep fade for one user may not be in deep fade for others. Therefore, a user may experience different performance with different codes and different time slots. In the previous work, a real time subcarrier allocation scheme is proposed to minimize the overall transmit power in [16]. In [17], the optimal code assignment scheme is proposed for MC-CDMA system. It has been shown that significant performance gain can be achieved by applying optimal code assignment. However, they didn't consider the time-sharing characteristic of channel. In order to further utilize the channel characteristics, we investigate the optimal assignment of codes with multiple time slots to optimize the system performance.

We consider MC-CDMA systems. We assume that there are N carrier frequencies and K simultaneous users in the system. The base station uses N orthogonal codes to create N orthogonal transmission channels. Thus we have a block of N channels with multiple time slots to serve the K users.

For the kth user, the assigned code for the transmission channel

$$\mathbf{c}_k = [c_{k,1}, c_{k,2}, \ldots, c_{k,N}]^T \tag{3.1}$$

determines the gain factors on the carriers. The base station sends the signal with the following complex representation

$$\sum_{n=1}^{N} c_{k,n} s_k(t) e^{j\omega_n t} \tag{3.2}$$

The structure of the MC transmitter for one user signal is depicted in Figure 3.1 [17]. We assume that the norm of each code is normalized to one unit.

In a certain time slot, the total transmitted signal is of the form

$$\sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n} s_k(t) e^{j\omega_n t} \tag{3.3}$$

We assume that the signals $s_k(t)$ are bandlimited, and are normalized to unit power. We also assume that the carrier frequencies are suitably chosen so that signals on different carriers are orthogonal and do not interfere with one another.

We now describe the physical channel model. We assume a frequency selective slow Rayleigh fading channel. By suitably choosing N and the

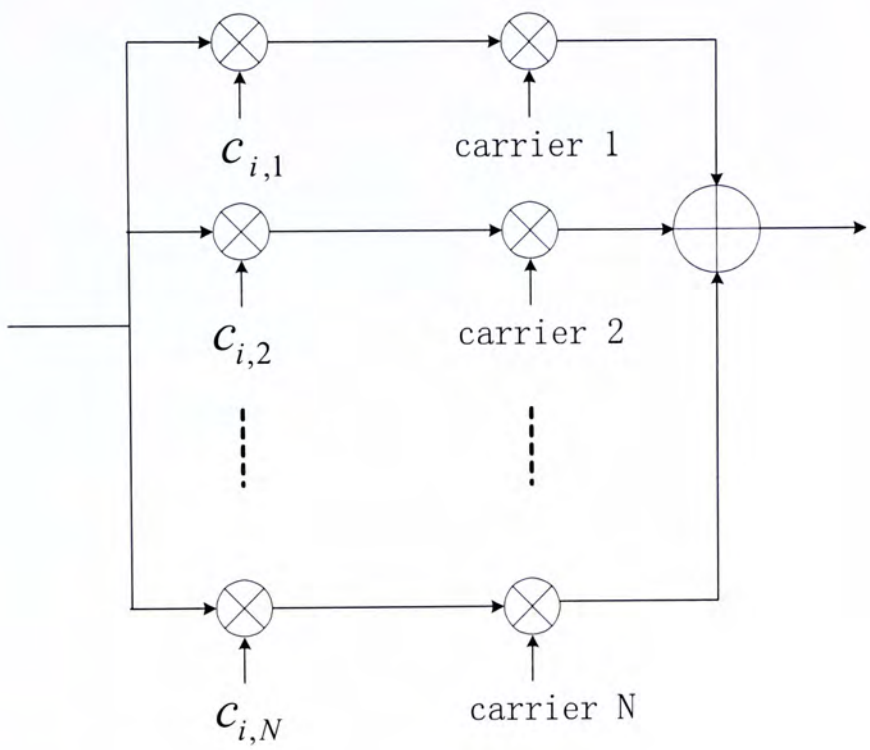

Figure 3.1: Block diagram of the MC transmitter for one signal

bandwidth of the narrowband signals, we can assume that each carrier undergoes independent frequency non-selective slow Rayleigh fading. We also assume the presence of additive white Gaussian noise (AWGN).

Without loss of generality, we consider the receiver for the first user. The received signal in the complex representation is given by

$$r_1(t) = g_1 \sum_{k=1}^{K} \sum_{n=1}^{N} c_{k,n} s_k(t) e^{j\omega_n t} \alpha_{1,n} + n_1(t) \tag{3.4}$$

where g_1 accounts for the large scale path loss, $\alpha_{1,n}$ accounts for the overall effects of phase shift and fading for the nth carrier of the first signal, and $n_1(t)$ represents AWGN. The fading coefficients are assumed to vary slowly so that their values can be accurately estimated.

The structure of the MC receiver is depicted in Figure 3.2 [17]. After low

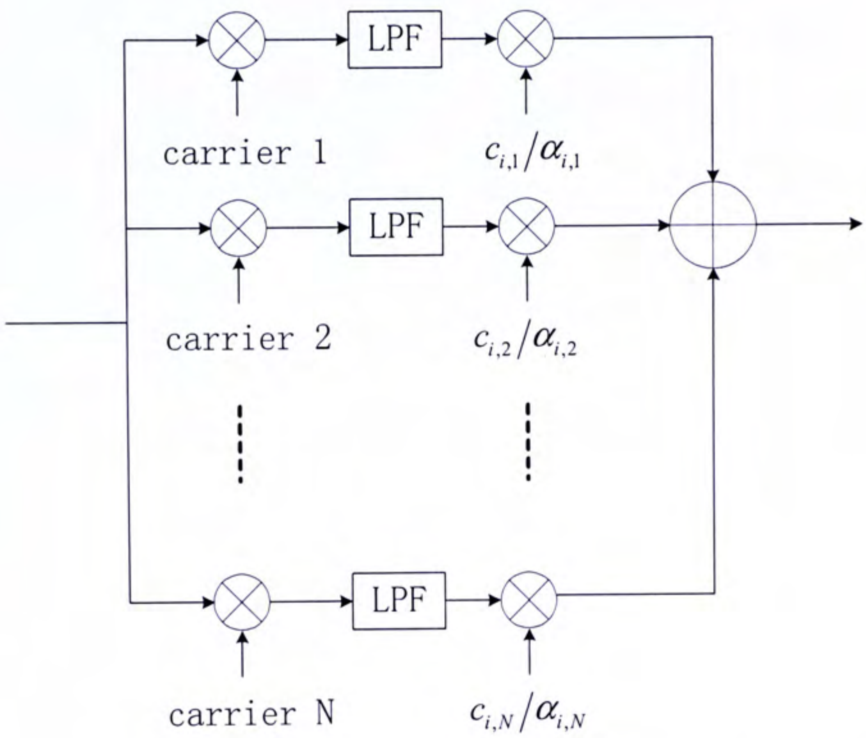

Figure 3.2: Block diagram of the MC receiver for one signal

pass filtering, the received vector signal is given by

$$\mathbf{r}_1(t) = g_1\mathbf{A}_1\left(\sum_{k=1}^{K} s_k(t)\mathbf{c}_k\right) + \mathbf{n}_1(t) \tag{3.5}$$

where $\mathbf{A}_1$ is a diagonal matrix whose nth element is $\alpha_{1,n}$.

With ideal channel state information (CSI), the receiver can compensate for the fading. The output of the receiver is therefore given by

$$z_1(t) = g_1 s_1(t) + g_1\sum_{k=2}^{K} s_k(t)\mathbf{c}_1^H\mathbf{c}_k + \mathbf{c}_1^H\mathbf{A}_1^{-1}\mathbf{n}_1(t) \tag{3.6}$$

For the orthogonal codes, the multiple access interference (MAI) accounted for by the middle term vanishes. The signal-to-noise ratio (SNR) is given by

$$\gamma_1 = \frac{g_1^2}{\sigma^2 W\mathbf{c}_1^H\mathbf{A}_1^{-1}\mathbf{A}_1^{-1H}\mathbf{c}_1} \tag{3.7}$$

where σ^2 is the variance of the AWGN contribution and W is the bandwidth of LPF at the receiver. For the codes which are not orthogonal, the cross correlation between any two codes may not be zero. The SNR is given by

$$\gamma_1 = \frac{g_1^2}{g_1^2 \sum_{k=2}^{K} \mathbf{c}_1^H \mathbf{c}_k \mathbf{c}_k^H \mathbf{c}_1 + \sigma^2 W \mathbf{c}_1^H \mathbf{A}_1^{-1} \mathbf{A}_1^{-1H} \mathbf{c}_1} \tag{3.8}$$

With suitable error control coding, the throughput can reach

$$r_1 = B \log_2(1 + \gamma_1) \tag{3.9}$$

where B is the system bandwidth.

In general, when nth code is assigned to kth user, the achievable data rate is

$$r_{k,n} = B \log_2(1 + \gamma_{k,n}) \tag{3.10}$$

where

$$\gamma_{k,n} = \frac{g_k^2}{\sigma^2 W \mathbf{c}_n^H \mathbf{A}_k^{-1} \mathbf{A}_k^{-1H} \mathbf{c}_n} \tag{3.11}$$

If we take into account the power allocation among users, then

$$r_{k,n} = B \log_2(1 + P_{k,n} \gamma_{k,n}) \tag{3.12}$$

Or we have

$$P_{k,n} = \frac{2^{r_{k,n}/B} - 1}{\gamma_{k,n}} \tag{3.13}$$

where $P_{k,n}$ is the allocated power to the kth user when he is assigned the nth code.

3.2 Optimization with Normalized Power

3.2.1 Transmission Time Minimization

N codes form N channels. The throughput of the kth user is $r_{k,n}$ when he is assigned the nth channel. All $r_{k,n}$'s can be found from the above equations. Suppose that each channel can be time shared by the users. As shown in Figure 3.3, the kth user is assigned $l_{k,n}$ time slots of the nth channel. Since each channel can transmit simultaneously, it is reasonable to define the transmission time as the maximum transmission time among the N channels, i.e.

$$T = \max_n \left\{ \sum_{k=1}^{K} l_{k,n} \cdot T_c \right\} \tag{3.14}$$

where T_c is the delay between consecutive time slots. For convenience, we set T_c to one hereafter.

We consider the case that the transmitter knows the amount of data bits to send for each user. Assume that the kth user has B_k bits to send and the channel keeps invariant during this interval. For minimization of the transmission time, the optimization problem can be formulated as follows:

$$\min \max_n \left\{ \sum_{k=1}^{K} l_{k,n} \right\} \tag{3.15}$$

subject to

$$\sum_{n=1}^{N} r_{k,n} l_{k,n} = B_k \text{ for } k = 1, \ldots, K$$

and

$$l_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

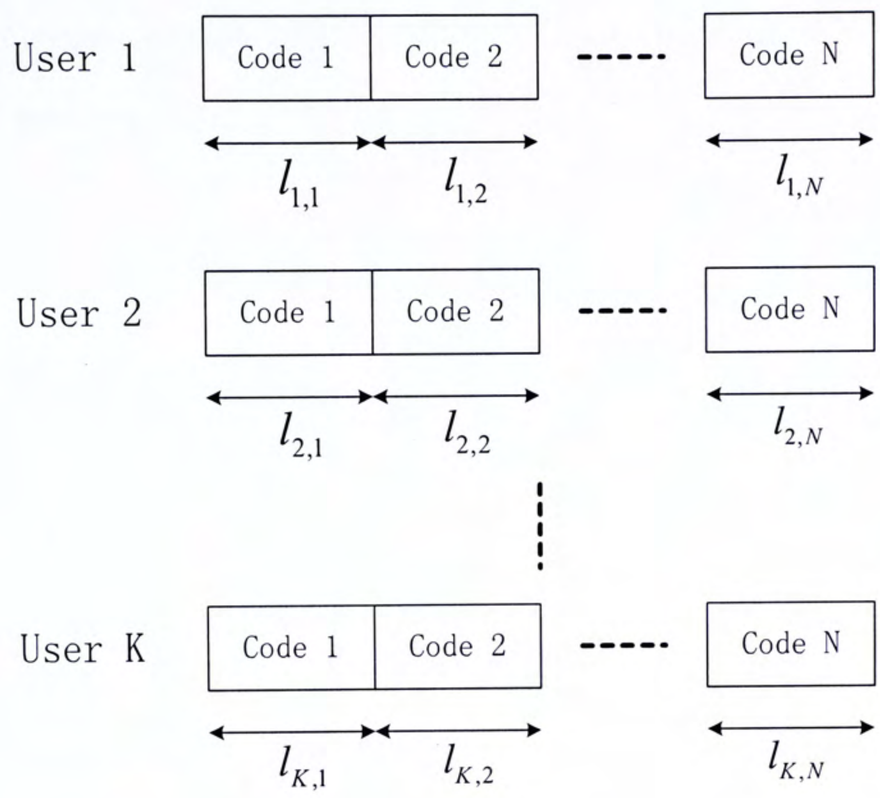

Figure 3.3: Time sharing of codes among K users

By introducing a new variable L, the original mini-max problem can be converted into a linear programming problem [18]-[20]:

$$\min L \tag{3.16}$$

subject to

$$\begin{aligned}
&\sum_{k=1}^{K} l_{k,n} - L \leq 0 \text{ for } n = 1, \ldots, N \\
&\sum_{n=1}^{N} r_{k,n} l_{k,n} = B_k \text{ for } k = 1, \ldots, K \\
&L \geq 0 \text{ and } l_{k,n} \geq 0 \text{ for all } k \text{ and } n.
\end{aligned}$$

Let $L_n = \sum_{k=1}^{K} l_{k,n}$ $(n = 1, 2, \ldots, N)$. If we want to minimize the maximum value among L_n's, an intuitive way is to make all L_n's take the same value. So the first N constrains may in fact be sharpened to equality, i.e., $\sum_{k=1}^{K} l_{k,n} = L$. Then the problem turns out to be:

$$\min L \tag{3.17}$$

subject to

$$\sum_{k=1}^{K} l_{k,n} = L \text{ for } n = 1, \ldots, N$$
$$\sum_{n=1}^{N} r_{k,n} l_{k,n} = B_k \text{ for } k = 1, \ldots, K$$
$$L \geq 0 \text{ and } l_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

The above three models are equivalent. In general, all the min-max problems with linear objective functions and linear constraints can be converted to linear programming. Solutions can be obtained readily with common mathematical software.

Here we assume that channel is invariant during the interval concerned. However, most wireless communication channels are time varying. When the users have large amount of data to send, the channel may not stay unchanged long enough and the assumption is not reasonable anymore. In this case, we need to update the optimal assignment periodically. By suitably choosing the update period, we can assume that the fading coefficients keep constant within the period.

By solving the above optimization problem, we will get real values instead of integer values for $l_{k,n}$. Thus we need to consider a practical suboptimal algorithm. In order to describe this algorithm, we need to introduce a concept first. Notice that each channel can be time shared by users, we can define *channel sharing factor* as follows:

Definition Channel Sharing Factor f is the maximum number of users who can share one channel in a transmitting period. Or we may say each channel is divided into f sub-channels in time domain.

By introducing Channel Sharing Factor, it seems that we have a total of $N \cdot f$ sub-channels to serve the users. Based on this, the algorithm can be described as a channel assignment method.

Algorithm:

1. Use an $N \cdot f$ dimensional flag vector to mark all $N \cdot f$ sub-channels unused.

2. Initialization: Each user chooses one good sub-channel which has not been used by other users.

3. Calculate each user's transmission time and find the user with maximum transmission time. Among the unused sub-channels, let him pick up the best one, i.e. with maximum throughput.

4. Update this user's transmission time and mark the chosen sub-channel used.

5. Go back to step 3 until all the sub-channels have been used.

Simulation results show that with suitable channel sharing factor f, the system performance will be very close to the optimal. Theoretically, let f go to infinity, i.e. we can arbitrarily divide each channel in the time domain, we will get the optimal solution via the algorithm.

If the data amount information is not available at the transmitter, we consider a time period of one normalized time slot during which the channel keeps invariant. The transmission time minimization problem can be formulated as

$$\min \max_{n} \sum_{k=1}^{K} t_{k,n} \tag{3.18}$$

subject to

$$\sum_{n=1}^{N} r_{k,n} t_{k,n} \geq R_k \text{ for all } k,$$

$$t_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

where $t_{k,n}$ is the time proportion of the nth channel being used by kth user and R_k is the required throughput of kth user.

By introducing a new variable, the above problem can be converted into a linear programming problem [20]

$$\min T \tag{3.19}$$

subject to

$$\sum_{k=1}^{K} t_{k,n} - T \leq 0 \text{ for all } n,$$

$$\sum_{n=1}^{N} r_{k,n} t_{k,n} \geq R_k \text{ for all } k,$$

$$t_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

3.2.2 Throughput Maximization

For a K-user system, the system performance is usually expressed as $U(r_1, \cdots, r_K)$, which is value of the *system utility function*. One commonly used definition of utility function is the summation of the individual user throughput

$$U(r_1, \cdots, r_K) = \sum_{k=1}^{K} r_k. \tag{3.20}$$

Transmitter designed to optimize such a utility function will result in the highest system capacity. However, users with poor channel condition could be discriminated (e.g., always get a zero value of throughput in the optimization process), and would suffer from starvation. This unfair situation is clearly undesirable [21], [22].

A classical resource sharing principle established to strike a balance between system capacity and fairness among users is the max-min fairness criterion, as discussed by Bertsekas and Gallager [23]. Max-min fairness has been applied to windows-based flow control, ABR service in ATM networks, wireless networks, and multicast protocols.

One possible way to define fairness for a resource allocation scheme is to require that each user should obtain the same transmission rate. For example, when K users access a single channel with capacity C, a fair allocation would give each user a transmission rate equal to C/K.

In the MC-CDMA system we considered, the throughput optimization problem can be formulated in max-min manner

$$\max \min_{k} \sum_{n=1}^{N} r_{k,n} t_{k,n} \tag{3.21}$$

subject to

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

We can use the similar maneuver as before to convert it into a linear programming problem [20]

$$\max R \tag{3.22}$$

subject to

$$\sum_{n=1}^{N} r_{k,n} t_{k,n} - R \geq 0 \text{ for all } k,$$

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

Max-min principle has a trend to guarantee the same data rate of all users. It is the simplest way to maximize the throughput and guarantee fairness among users at the same time. In fact, we can extend it by introducing fairness factor a_k of kth user. Then, the objective would be

$$\max \min_{k} \sum_{n=1}^{N} r_{k,n} * t_{k,n} / a_k \tag{3.23}$$

3.2.3 Performance

In this subsection, we investigate the performance of optimal and suboptimal algorithms when the data amount information is known at the transmitter. We consider the case where $N = 16$. Assume that users are uniformly distributed in a circular cell of radius of 2 units. The gain factor for each user

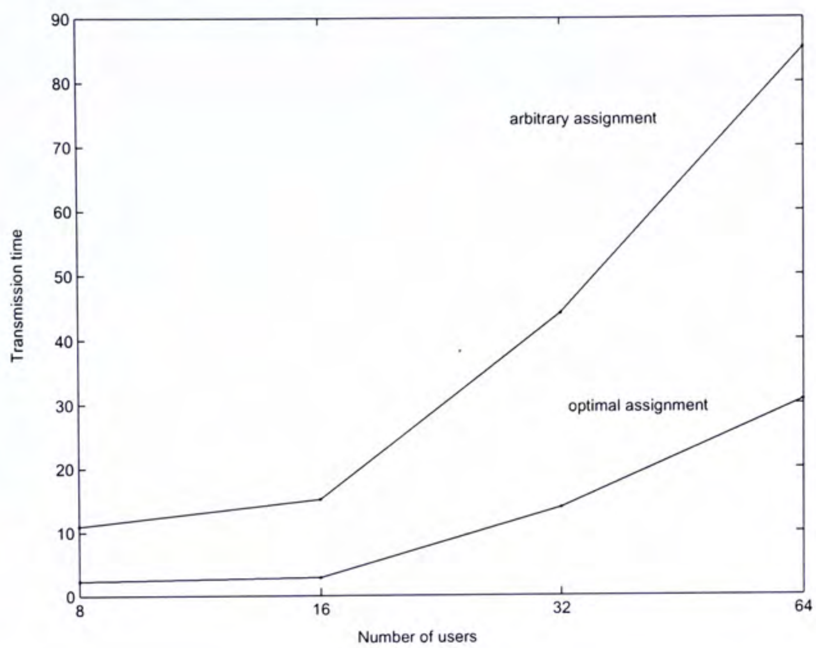

Figure 3.4: Performance of arbitrary and optimal assignments

is inversely proportional to the square of the distance from the base station so that the power of the signal decays as the fourth power of the distance. The fading coefficients $\alpha_{k,n}$'s are independently and identically distributed (iid) complex Gaussian random variables of zero mean and unit variance. The variance of the AWGN contribution is normalized to one unit. The data length R_i's are uniformly distributed from 0 to 10.

We plot the transmission time (the maximum transmission time among the N channels) against the number of users. The system performance of arbitrary and optimal assignments are compared in Figure 3.4. We observe that system performance can be tremendously improved with our optimization scheme.

In Figure 3.5, we investigate the performance of suboptimal algorithms

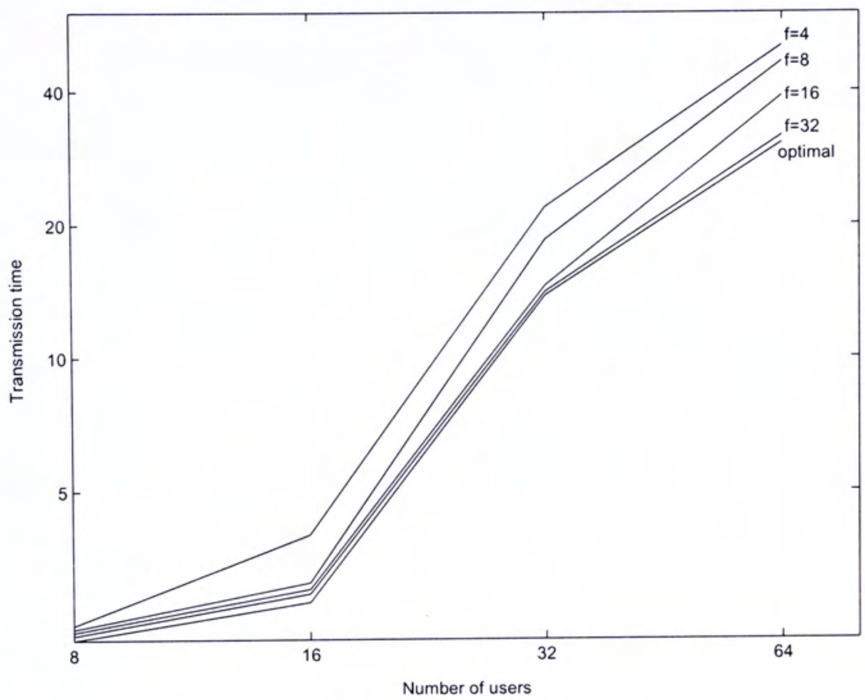

Figure 3.5: Performance of optimal assignment and suboptimal algorithm with Channel Sharing Factor $f = 4, 8, 16, 32$

Table 3.1: Near optimal Channel Sharing Factor for different number of users ($N = 16$)

Number of users (K)	8	16	32	64
Channel sharing factor (f)	4	8	16	32

with different values of channel sharing factor f. It shows that for 8 users, the performance is very close to the optimal with $f = 4$. While for 16 users, we need $f = 8$ to obtain near optimal performance. We list the observations for different number of users in Table I.

Further simulation results give an empirical formula of near optimal channel sharing factor:

$$f_0 = \frac{8 \cdot K}{N} \tag{3.24}$$

When f equals f_0, we can obtain near optimal solution via the proposed algorithm. From (3.24), we note that $f_0 = 1$ when $N = 8K$, i.e., it is not necessary to perform time-sharing of channel when we have enough number of channels. In fact, when $N > 4K$, system performance with only code assignment is good enough.

3.3 Mathematical Programming

Before further discussing the optimization problem with power constraints, we introduce some concepts concerning mathematical programming.

A mathematical programming problem, or optimization problem, has the following form [20]

$$\min f_0(x) \tag{3.25}$$

subject to

$$f_i(x) \leq 0, i = 1, \cdots, m$$

$$h_i(x) = 0, i = 1, \cdots, p$$

where the vector $x = [x_1, x_2, \cdots, x_n]$ is the *optimization variable* of the problem, the function $f_0 : R^n \rightarrow R$ is called the *objective function* or *cost function*, the functions $f_i : R^n \rightarrow R, i = 1, \cdots, m$, are called the *inequality constraint functions*, and the corresponding equality $f_i(x) \leq 0, i = 1, \cdots, m$, are called *inequality constraints*. The functions $h_i : R^n \rightarrow R, i = 1, \cdots, p$, are called the *equality constraint functions*, and the corresponding equality $h_i(x) = 0, i = 1, \cdots, p$, are called *equality constraints*. If there are no constraints (i.e. $m = p = 0$), the problem is said to be *unconstrained*. A vector x^* is called *optimal solution* of the above programming problem, if it has the smallest objective value among all vectors that satisfy the constraints: for any z with $f_1(z) \leq 0, \cdots, f_m(z) \leq 0$ and $h_1(z) = 0, \cdots, h_p(z) = 0$, we have $f_0(z) \geq f_0(x^*)$.

Generally, we consider families or classes of optimization problems, which are characterized by different forms of objective and constraint functions. One important example is the *linear programming* problem, whose objective and constraint functions $f_0, \cdots, f_m, h_1, \cdots, h_p$ are linear, i.e. satisfy $f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y)$ for all $x, y \in R^n$ and all $\alpha, \beta \in R$. If the optimization problem is not linear, it is called a *nonlinear programming* problem [20].

3.3.1 Nonlinear Programming

Usually, nonlinear programming is the term used to describe an optimization problem where the objective or constraint functions are not linear, but not known to be convex. Sadly, there are no effective methods for solving the general nonlinear programming problem. Even simple looking problems with as few as ten variables can be extremely challenging, while problems with a few hundreds of variables can be intractable. Methods for the general nonlinear programming problem therefore take several different approaches, each of which involves some compromise.

One class of approaches is called *local optimization* [20]. In this kind of approaches, the compromise is to give up seeking the optimal x, which minimizes the objective over all feasible points. Instead we seek a point that is only locally optimal, which means that it minimizes the objective function among all feasible points that are near it, but is not guaranteed to have a lower objective value than all other feasible points. A large fraction of the research on general nonlinear programming has focused on methods for local optimization, which as a consequence are well developed.

Local optimization methods can be fast, can handle large-scale problems, and are widely applicable, since they only require differentiability of the objective and constraint functions. As a result, local optimization methods are widely used in applications where there is value in finding a good point, if not the very best. In an engineering design application, for example, local optimization can be used to improve the performance of a design originally obtained by manual, or other, design methods.

There are several disadvantages of local optimization methods, beyond (possibly) not finding the true, globally optimal solution. The methods require an initial guess for the optimization variable. This initial guess or starting point is critical, and can greatly affect the objective value of the local solution obtained. Little information is provided about how far from (globally) optimal the local solution is [20]. Local optimization methods are often sensitive to algorithm parameter values, which may need to be adjusted for a particular problem, or family of problems.

Another family of approaches for nonlinear programming is *global optimization* approach [20]. In global optimization, the true global solution of the optimization problem is found; the compromise is efficiency. The worst-case complexity of global optimization methods grows exponentially with the problem sizes, i.e., the dimension of the x vector and the number of constraints; the hope is that in practice, for the particular problem instances encountered, the method is far faster. While this favorable situation does occur, it is not typical. Even small problems, with a few tens of variables, can take a very long time (e.g., hours or days) to solve. Usually, global optimization is used for problems with a small number of variables, where computing time is not critical, and the value of finding the true global solution is very high.

3.3.2 Convex Programming

In point of view of convexity, we can also divide the programming problems into two classes, *convex* and *nonconvex* programming. A convex optimization problem is one of the form [20]

$$\min f_0(x) \tag{3.26}$$

subject to

$$f_i(x) \leq 0, i = 1, \cdots, m$$

$$a_i^T x = b_i, i = 1, \cdots, p$$

where $f_0, \cdots, f_m$ are convex functions, i.e., satisfy

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \tag{3.27}$$

for all $x, y \in R^n$ and all $\alpha, \beta \in R$ with $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$.

Comparing with the general standard form problem, the convex problem has three additional requirements:

- The objective function must be convex;
- The inequality constraint functions must be convex;
- The equality constraint functions $h_i(x) = a_i^T x - b_i$ must be affine.

There is in general no analytical formula for the solution of convex optimization problems, but (as with linear programming problems) there are very effective methods for solving them [20]. Interior-point methods work very well in practice, and in some cases can be proved to solve the problem to a specified accuracy with a number of operations that does not exceed a polynomial of the problem dimensions.

Using convex optimization is, at least conceptually, very much like using linear programming. If we can formulate a problem as a convex optimization problem, then we can solve it efficiently, just as we can solve a linear problem efficiently. With only a bit of exaggeration, we can say that, if you formulate a practical problem as a convex optimization problem, then you have solved the original problem. Up to now, many efficient algorithms to solve this kind of problem have already been proposed by many researchers [20]. Among those algorithms, interior point method has been proved to be effective and efficient for convex programming. At the same time, many software programs based on these algorithms are developed by research groups in universities and software companies. Currently, the popular solvers are MOSEK, MINOS, KNITRO, which are commercial software, and SOLNP, which is a free solver.

A very important property of convex programming problem is that any local optimal solution is also the global optimal solution [20]. To see this, suppose that x is locally optimal for a convex programming problem, i.e., x is feasible and

$$f_0(x) = \inf\{f_0(z) \mid z \text{ is feasible, } \|z - x\|_2 \leq R\} \tag{3.28}$$

for some $R \geq 0$. Now, if we suppose that x is not globally optimal, that is, there is another feasible solution y such that $f_0(y) < f_0(x)$. Evidently $\|y - x\|_2 > R$, since otherwise $f_0(x) < f_0(y)$. We consider the point z of the form

$$z = (1-\theta)x + \theta y, \quad \theta = \frac{R}{2\|y - x\|_2} \tag{3.29}$$

Then we have $\|z - x\|_2 = R/2 < R$. By the convexity of $f_0(x)$, we have

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y) < f_0(x) \tag{3.30}$$

which contradicts (3.16). Therefore, there is no feasible y with $f_0(y) < f_0(x)$, that is, x is globally optimal. Hence, we can conclude that the local optimal solution of a convex nonlinear programming problem is also the global optimal.

3.4 Optimization with Power Allocation

In section 3.2, for simplicity, we model the optimization problem without considering power allocation. However, power budget is one key concern in wireless communication system implementation. In this section, we will re-formulate the optimization problem by introducing power budget constraint.

3.4.1 Transmission Time Minimization with Power Allocation

We consider the general version of problem (3.19) with power allocation. By substituting equation (3.12), the transmission time minimization problem can be formulated as

$$\min \max_{n} \sum_{k=1}^{K} t_{k,n} \tag{3.31}$$

subject to

$$\sum_{k=1}^{K} \sum_{n=1}^{N} P_{k,n} t_{k,n} \leq P_{Budget},$$

$$\sum_{n=1}^{N} B \log_2(1 + P_{k,n}\gamma_{k,n}) \cdot t_{k,n} \geq R_k \text{ for all } k,$$

$$t_{k,n} \geq 0 \text{ and } P_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

This is a nonlinear min-max problem. However, it is not intractable. Via the help of variable replacement, we can convert it into convex programming

[24]. Let $Q_{k,n} = P_{k,n}t_{k,n}$, then the above problem can be reformulated as following

$$\min T \tag{3.32}$$

subject to

$$\sum_{k=1}^{K} t_{k,n} - T \leq 0 \text{ for all } n,$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} Q_{k,n} \leq P_{Budget},$$

$$\sum_{n=1}^{N} B\log_2(1+\frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n} \geq R_k \text{ for all } k,$$

$$t_{k,n} \geq 0 \text{ and } Q_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

It is not difficult to prove this is a convex programming problem (see Appendix A). As discussed in the previous section, it can be solved efficiently by the well developed methods, e.g., the interior point algorithm [20].

3.4.2 Throughput Maximization with Power Allocation

For throughput, we follow the max-min principle again as in (3.22). By adding the power budget constraints and substituting equation (3.12), the throughput maximization problem can be formulated as

$$\max\min_{k}\sum_{n=1}^{N} B\log_2(1+P_{k,n}\gamma_{k,n}) \cdot t_{k,n} \tag{3.33}$$

subject to

$$\sum_{k=1}^{K}\sum_{n=1}^{N} P_{k,n}t_{k,n} \leq P_{Budget},$$

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ and } P_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

Similarly, let $Q_{k,n} = P_{k,n} t_{k,n}$, the problem can be converted into

$$\max R \tag{3.34}$$

subject to

$$\sum_{n=1}^{N} B \log_2(1 + \frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n} - R \geq 0 \text{ for all } k,$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} Q_{k,n} \leq P_{Budget},$$

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ and } Q_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

Notice that the objective function and constraints of the above problem are very similar to (3.32). We can easily verify that it is convex programming.

3.4.3 Power Minimization

As power is being taken into consideration, we can model a power minimization problem while treating $r_{k,n}$'s and $t_{k,n}$'s as optimization variables. By substituting equation (3.13), we have

$$\min \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{2^{r_{k,n}/B} - 1}{\gamma_{k,n}} \cdot t_{k,n} \tag{3.35}$$

subject to

$$\sum_{n=1}^{N} r_{k,n} t_{k,n} \geq R_k \text{ for all } k,$$

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ and } r_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

Let $s_{k,n} = r_{k,n} t_{k,n}$, the above problem can be converted into convex programming [24]

$$\min \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{2^{s_{k,n}/Bt_{k,n}} - 1}{\gamma_{k,n}} \cdot t_{k,n} \tag{3.36}$$

subject to

$$\sum_{n=1}^{N} s_{k,n} \geq R_k \text{ for all } k,$$

$$\sum_{k=1}^{K} t_{k,n} \leq 1 \text{ for all } n,$$

$$t_{k,n} \geq 0 \text{ and } s_{k,n} \geq 0 \text{ for all } k \text{ and } n.$$

The convexity verification of the above problem can be found in Appendix B.

3.5 Long-range Optimization

Previously, we only consider one normalized time slot. It would be more meaningful to solve the long-range optimization problem if given a longer period of time.

3.5.1 Long-range Transmission Time Minimization

The problem can be formulated as [24]

$$\min \max_{n} \sum_{l=1}^{L} \sum_{k=1}^{K} t_{k,n,l} \tag{3.37}$$

subject to

$$\sum_{k=1}^{K} \sum_{n=1}^{N} P_{k,n,l} t_{k,n,l} \leq P_{Budget} \text{ for all } l,$$

$$\sum_{l=1}^{L} \sum_{n=1}^{N} B \log_2(1 + P_{k,n,l}\gamma_{k,n,l}) \cdot t_{k,n,l} \geq R_k L \text{ for all } k,$$

$$t_{k,n,l} \geq 0 \text{ and } P_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

We propose two methods to solve the problem, and will compare their performance later.

1. *Divide the period into several normalized time slots, and via previous solution, perform optimization in each time slot one by one.*

In the lth time slot, we try to solve this problem

$$\min \max_{n} \sum_{k=1}^{K} t_{k,n,l} \tag{3.38}$$

subject to

$$\sum_{k=1}^{K} \sum_{n=1}^{N} P_{k,n,l} t_{k,n,l} \leq P_{Budget},$$

$$\sum_{n=1}^{N} B \log_2(1 + P_{k,n,l}\gamma_{k,n,l}) \cdot t_{k,n,l} \geq R_k \text{ for all } k,$$

$$t_{k,n,l} \geq 0 \text{ and } P_{k,n,l} \geq 0 \text{ for all } k \text{ and } n.$$

We can solve the above problem by converting it into convex programming. However, note that the above problem is not equivalent to (3.16), since we treat each time slot in an independent manner. Thus, the solution is not optimal, but near to the optimal.

For the fading coefficient estimation in each time slot, we prefer a short-range prediction method with low complexity. Polynomial fitting prediction based on several past CSI samples should be the one we are looking for [25]. Assume M past samples $\{\alpha(n-i)\}_{i=0}^{M-1}$ are available and the sampling period is T. We use a polynomial $A(t)$ with degree $M-1$ to fit the samples.

$$A(t) = \sum_{i=0}^{M-1} b_i t^i \tag{3.39}$$

should satisfy

$$A(t) = \begin{cases} \alpha(n) & \text{if } t = nT \\ \alpha(n-1) & \text{if } t = (n-1)T \\ \alpha(n-2) & \text{if } t = (n-2)T \\ \vdots & \vdots \\ \alpha(n-M+1) & \text{if } t = (n-M+1)T \end{cases} \tag{3.40}$$

All the polynomial coefficients can be found by solving the following set of linear equations

$$\mathbf{Cb} = \mathbf{a} \tag{3.41}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & nT & \cdots & (nT)^{M-1} \\ 1 & (n-1)T & \cdots & ((n-1)T)^{M-1} \\ 1 & (n-2)T & \cdots & ((n-2)T)^{M-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & (n-M+1)T & \cdots & ((n-M+1)T)^{M-1} \end{bmatrix} \tag{3.42}$$

$$\mathbf{a} = [\ \alpha(n)\ \alpha(n-1)\ \alpha(n-2)\ \cdots\ \alpha(n-M+1)\ ]^T \tag{3.43}$$

with unknown variables

$$\mathbf{b} = [\ b_0\ b_1\ b_2\ \cdots\ b_{M-1}\ ]^T \tag{3.44}$$

Then we can calculate the predicted value of the next CSI $\alpha(n+1)$.

$$\alpha(n+1) = \sum_{i=0}^{M-1} b_i((n+1)T)^i \tag{3.45}$$

Notice that $\mathbf{C}$ is Vandermonde matrix. Based on its property, we know that the calculation of $\alpha(n+1)$ should be independent of the value n. So we can choose $\mathbf{C}$ as

$$\mathbf{C} = \begin{bmatrix} 1 & M & \cdots & M^{M-1} \\ 1 & M-1 & \cdots & (M-1)^{M-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 2 & \cdots & 2^{M-1} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{3.46}$$

Then calculate $\mathbf{b}$ by solving

$$\mathbf{Cb} = \mathbf{a} \tag{3.47}$$

The calculation of $\alpha(n+1)$ can be simplified as

$$\alpha(n+1) = \sum_{i=0}^{M-1} b_i (M+1)^i \tag{3.48}$$

2: *Use long-range channel prediction method to predict the CSI during the period in advance and perform optimization as a whole.*

We use a long-range channel prediction method which employs an autoregressive (AR) model to characterize the fading channel [26]-[28]. If we can predict far enough, we can then solve the original optimization problem.

Let $Q_{k,n,l} = P_{k,n,l} t_{k,n,l}$. After conversion, we obtain

$$\min T \tag{3.49}$$

subject to

$$\sum_{l=1}^{L}\sum_{k=1}^{K} t_{k,n,l} - T \leq 0 \text{ for all } n,$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} Q_{k,n,l} \leq P_{Budget} \text{ for all } l,$$

$$\sum_{l=1}^{L}\sum_{n=1}^{N} B\log_2(1 + \frac{Q_{k,n,l}\gamma_{k,n,l}}{t_{k,n,l}}) \cdot t_{k,n,l} \geq R_k L \text{ for all } k,$$

$$t_{k,n,l} \geq 0 \text{ and } Q_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

Since the predicted value of the channel coefficients are available there, we can calculate all $\gamma_{k,n,l}$'s beforehand. The remaining work is to solve the above convex programming problem.

3.5.2 Long-range Throughput Maximization

The long period throughput maximization problem can be formulated as

$$\max \min_{k} \sum_{l=1}^{L} \sum_{n=1}^{N} B \log_2(1 + P_{k,n,l}\gamma_{k,n,l}) \cdot t_{k,n,l} \tag{3.50}$$

subject to

$$\sum_{k=1}^{K} \sum_{n=1}^{N} P_{k,n,l} t_{k,n,l} \leq P_{Budget} \text{ for all } l,$$

$$\sum_{k=1}^{K} t_{k,n,l} \leq 1 \text{ for all } n \text{ and } l,$$

$$t_{k,n,l} \geq 0 \text{ and } P_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

We can employ similar methods to solve this problem as before. For the case of long-range channel estimation, let $Q_{k,n,l} = P_{k,n,l} t_{k,n,l}$, the problem can be converted into convex programming

$$\max R \tag{3.51}$$

subject to

$$\sum_{l=1}^{L} \sum_{n=1}^{N} B \log_2(1 + \frac{Q_{k,n,l}\gamma_{k,n,l}}{t_{k,n,l}}) \cdot t_{k,n,l} - R \geq 0 \text{ for all } k,$$

$$\sum_{k=1}^{K} \sum_{n=1}^{N} Q_{k,n,l} \leq P_{Budget} \text{ for all } l,$$

$$\sum_{k=1}^{K} t_{k,n,l} \leq 1 \text{ for all } n \text{ and } l,$$

$$t_{k,n,l} \geq 0 \text{ and } Q_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

3.5.3 Long-range Power Minimization

The long period power minimization problem can be formulated as [24]

$$\min \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{2^{r_{k,n,l}/B} - 1}{\gamma_{k,n,l}} \cdot t_{k,n,l} \tag{3.52}$$

subject to

$$\sum_{l=1}^{L} \sum_{n=1}^{N} r_{k,n,l} t_{k,n,l} \geq R_k L \text{ for all } k,$$

$$\sum_{k=1}^{K} t_{k,n,l} \leq 1 \text{ for all } n \text{ and } l,$$

$$t_{k,n,l} \geq 0 \text{ and } r_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

For the case of long-range channel estimation, we can convert the above problem into convex programming. Let $s_{k,n,l} = r_{k,n,l} t_{k,n,l}$, we have

$$\min \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{2^{s_{k,n,l}/Bt_{k,n,l}} - 1}{\gamma_{k,n,l}} \cdot t_{k,n,l} \tag{3.53}$$

subject to

$$\sum_{l=1}^{L} \sum_{n=1}^{N} s_{k,n,l} \geq R_k L \text{ for all } k,$$

$$\sum_{k=1}^{K} t_{k,n,l} \leq 1 \text{ for all } n \text{ and } l,$$

$$t_{k,n,l} \geq 0 \text{ and } s_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

Here we propose another sub-optimal real-time assignment method with low computational complexity [24]. The purpose is to simplify the original problem into an assignment problem and perform assignment in each time slot.

We consider the lth time slot, the problem turns out to be

$$\min \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{2^{r_{k,n,l}/B} - 1}{\gamma_{k,n,l}} \cdot t_{k,n,l} \tag{3.54}$$

subject to

$$\sum_{n=1}^{N} r_{k,n,l} t_{k,n,l} \geq R_k \text{ for all } k \text{ and } l,$$

$$\sum_{k=1}^{K} t_{k,n,l} \leq 1 \text{ for all } n \text{ and } l,$$

$$t_{k,n,l} \geq 0 \text{ and } r_{k,n,l} \geq 0 \text{ for all } k, n \text{ and } l.$$

Note the rate requirement constraints in the above problem. For the minimization problem, they can be sharpened to equalities

$$\sum_{n=1}^{N} r_{k,n,l} t_{k,n,l} = R_k \text{ for all } k \text{ and } l, \tag{3.55}$$

We try to give the series of rate variables constant value, i.e., let $r_{k,n,l} = \tilde{R}_k = KR_k/N$. The optimization problem can be simplified to an assignment problem [24]

$$\min \sum_{k=1}^{K} w_{k,l} \sum_{n=1}^{N} \frac{t_{k,n,l}(2^{\tilde{R}_k/B} - 1)}{\gamma_{k,n,l}} \tag{3.56}$$

subject to

$$\sum_{k=1}^{K} t_{k,n,l} = 1 \text{ for all } k \text{ and } l,$$

$$t_{k,n,l} = 0 \text{ or } 1 \text{ for all } k, n \text{ and } l.$$

where $w_{k,l}$'s are factors to adjust the weight of each user to guarantee the fair resource share among users through the whole period. We choose $w_{k,l}$ as

$$w_{k,l} = \frac{\sum_{j=1}^{l} \sum_{n=1}^{N} t_{k,n,j}}{NL/K - \sum_{j=1}^{l} \sum_{n=1}^{N} t_{k,n,j}} \tag{3.57}$$

3.5.4 Performance

The performance evaluation is given in this part. We compare the performance of the optimization schemes with two different channel estimation methods.

We consider the case that the number of codes $N = 16$. We assume that users are uniformly located in a 120 degree sector with cell radius of 1 km. For the mobile wireless channel, the large scale path loss and the fading coefficients are generated by Hata model and Clarke model [29] respectively. We also assume the variance of the AWGN contribution is normalized to one unit. The code sets, unless otherwise stated, are the random orthogonal codes. In our simulations, we investigate two speeds of user mobility, more specifically, 3.6 km/h and 72 km/h, which are typical pedestrian speed and driving speed corresponding to slow and fast fading, respectively. Other parameters used in the simulations are listed as follows:

- Number of multipaths: 20
- Transmitter antenna height: 30 m
- Receiver antenna height: 1.5 m
- RF frequency : 900 MHz

In Figure 3.6 and Figure 3.7, we compare transmission time of the schemes with two different channel estimation methods and the one with ideal channel state information in slow and fast fading channels respectively. We find that

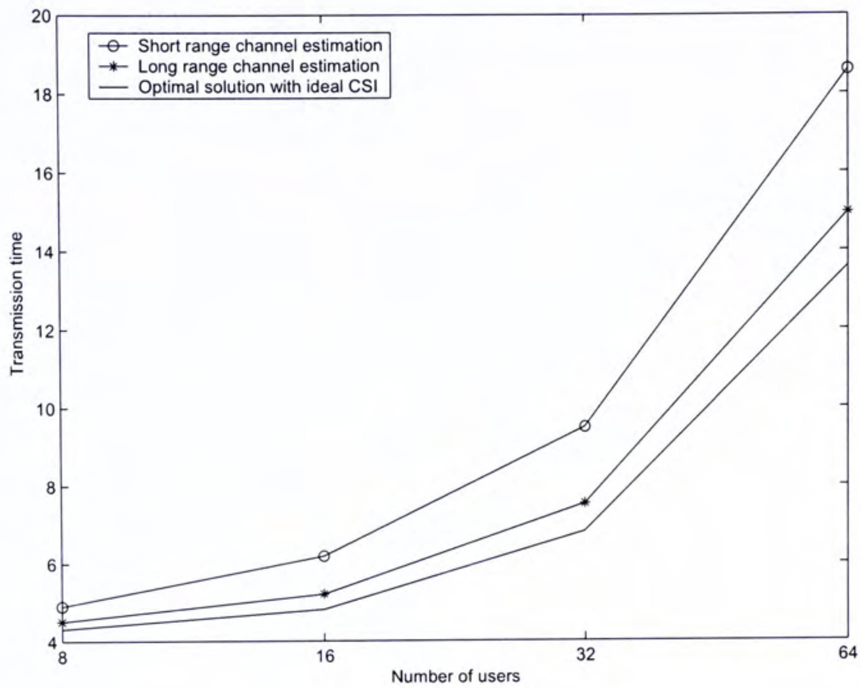

Figure 3.6: Transmission time against different number of users for slow fading channels

the one with long range channel estimation always outperforms short range channel estimation in slow fading channels. However, this is not always the case for fast fading. The reason is that slow fading channel is easier to estimate and the estimation is more accurate. It is straightforward to see that long range channel estimation will perform better. Figure 3.8 and Figure 3.9 show the similar results for throughput against the number of users.

For the power, we plot one more curve of our real-time assignment method. From Figure 3.10 and Figure 3.11, we can see that the real-time assignment method can be used as a good replacement of channel estimation methods in both slow and fast fading channels.

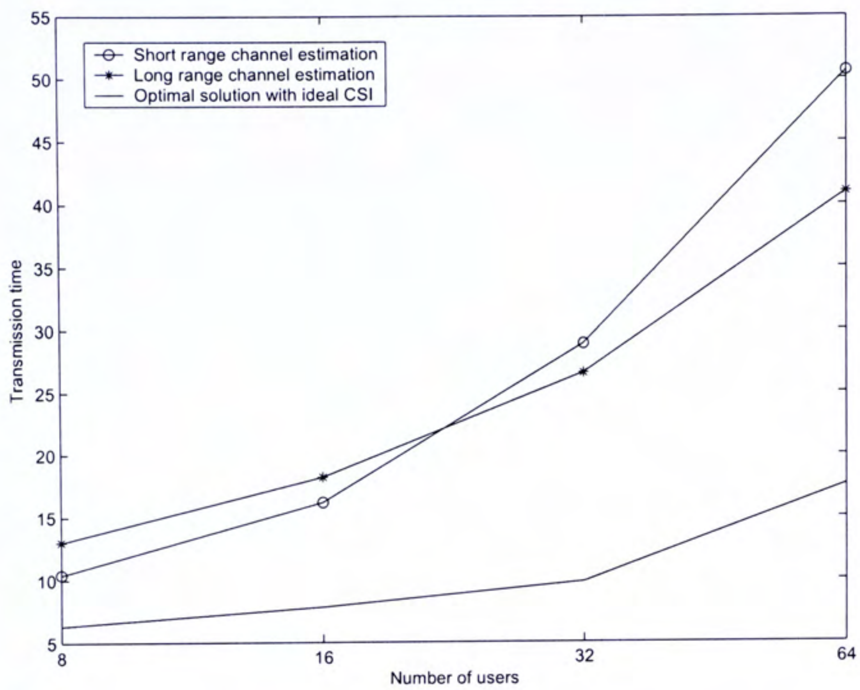

Figure 3.7: Transmission time against different number of users for fast fading channels

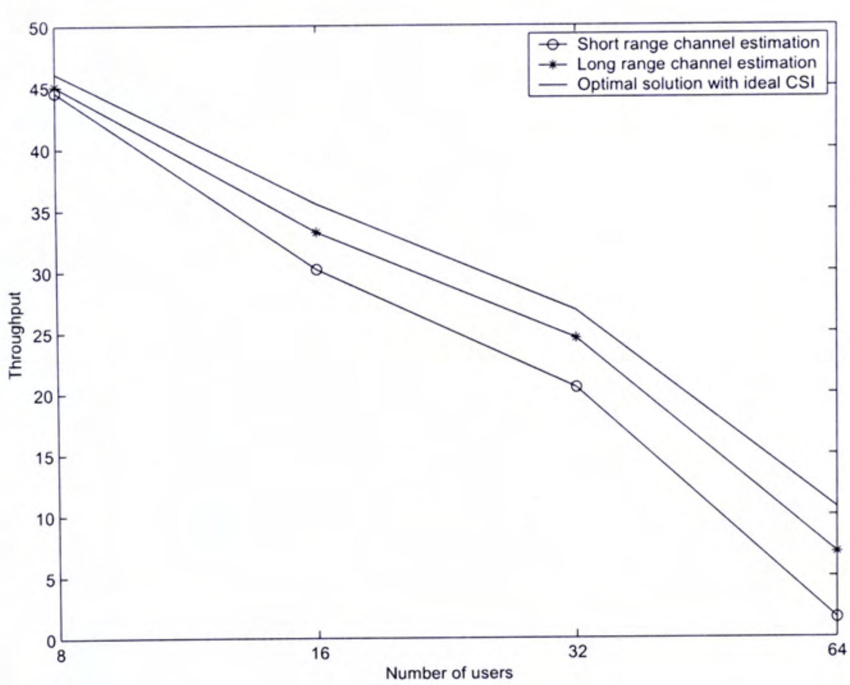

Figure 3.8: Throughput against different number of users for slow fading channels

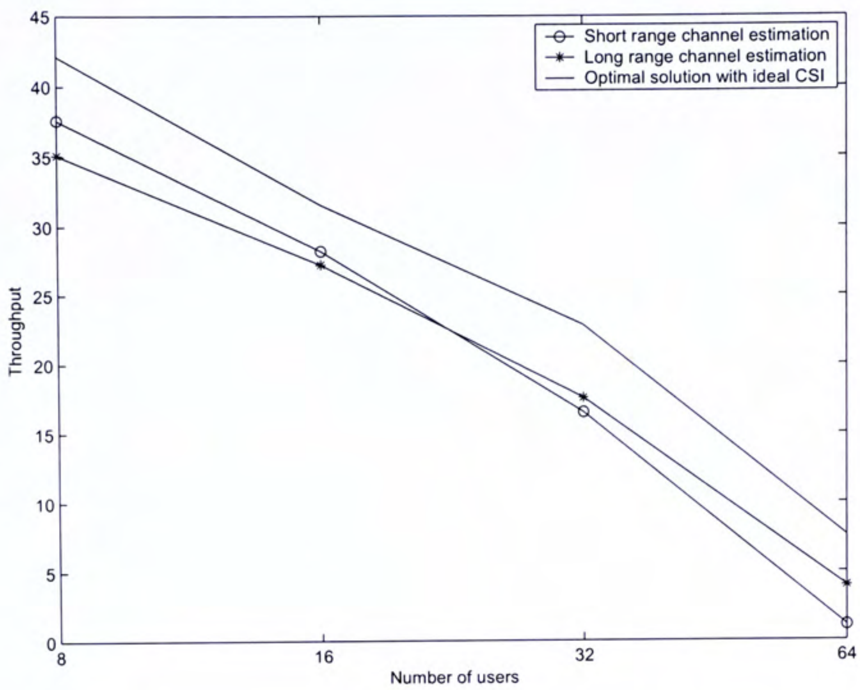

Figure 3.9: Throughput against different number of users for fast fading channels

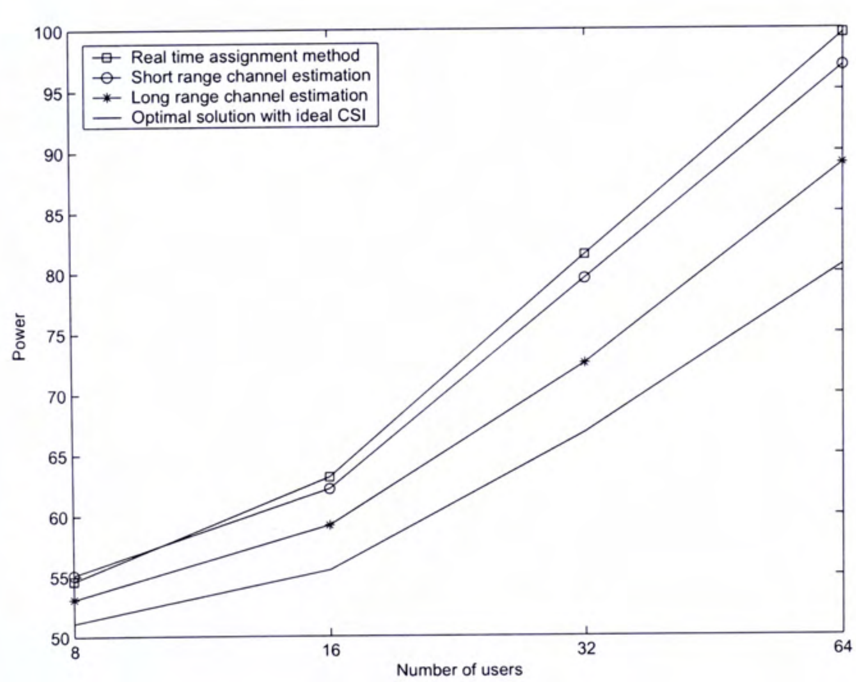

Figure 3.10: Required power against different number of users for slow fading channels

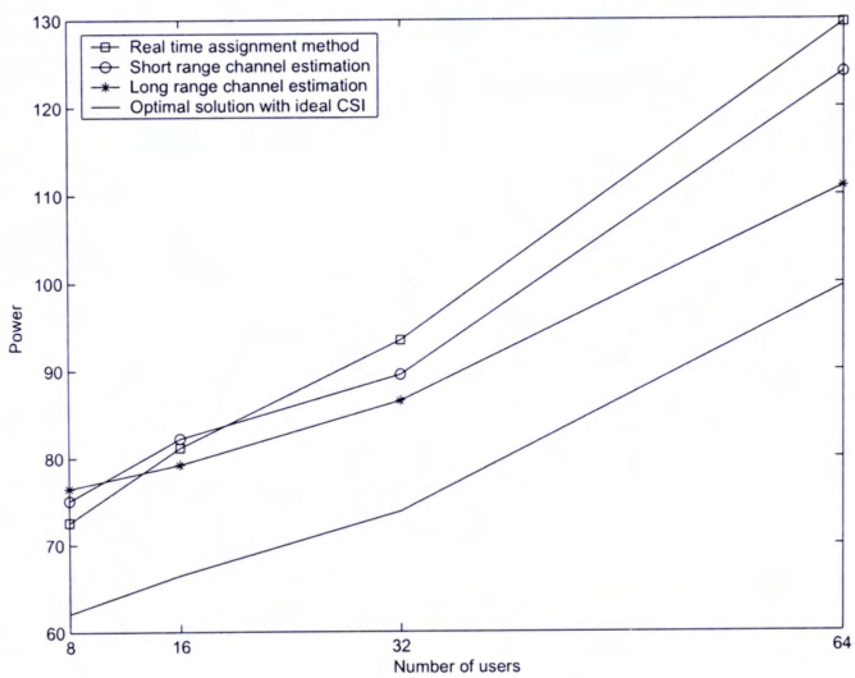

Figure 3.11: Required power against different number of users for fast fading channels

Chapter 4

Joint Scheduling and Resource Allocation

4.1 Queueing Model

In this chapter, we use a queueing model to formulate the long term performance information of users. We consider the downlink data transmission of MC-CDMA systems for delay-sensitive traffic. The purpose is to minimize the queueing delay by assigning system resource dynamically according to the queue and channel state information. For delay-sensitive traffic, QoS requirement can be expressed in terms of the delay violation probability [30]

$$Pr\{T > T_{max}\} \leq \epsilon \tag{4.1}$$

where T is a packet waiting time, and T_{max} and ϵ are the delay bound and the maximum acceptable probability respectively. According to [31], when the traffic load is heavy, the following approximation can be used in $G/G/1$ queues

$$Pr\{T > T_{max}\} \approx \exp\{-\frac{T_{max}}{W}\} \tag{4.2}$$

where W is the expectation of T, $E[T]$. Equation (4.2) implies that shortening the mean waiting time W is approximately equivalent to minimizing the delay violation probability in the heavy-traffic scenario. Since most delay violations occur during heavy traffic load, it is reasonable to use the mean waiting time to formulate the problems [30].

Networks users' performance can be modeled through *utility function.* It is shown in [32] that maximizing utility is able to balance resource efficiency and fairness. Here, we express the system utility as a function of the mean waiting time.

We consider the same MC-CDMA system as before. For the case of orthogonal codes and the signals are normalized to unit power, the SNR is given by

$$\gamma_k[n] = \frac{g_k^2}{\sigma^2 W \mathbf{c}_n^H \mathbf{A}_k^{-1} \mathbf{A}_k^{-1H} \mathbf{c}_n} \tag{4.3}$$

We use $\gamma_k[n,l]$ to represent SNR at time slot l. When the nth code is assigned to the kth user, the achievable data rate during time slot l is

$$r_k[n,l] = B\log_2(1 + \gamma_k[n,l])$$

If we take into account the power allocation among users, then

$$r_k[n,l] = B\log_2(1 + P_k[n,l]\gamma_k[n,l]) \tag{4.4}$$

The total data rate of user k is

$$r_k[l] = \sum_{n=1}^{N} \rho_k[n,l] r_k[n,l] \tag{4.5}$$

where $\rho_k[n,l]$ can be interpreted as the sharing factor of user k to code n at time slot l.

Assume that user k is associated with a mean waiting time W_k, and the corresponding utility is $U_k(W_k)$. Obviously, with longer delay, the user is less satisfied. So $U_k(W_k)$ should be decreasing and concave, which implies that $\frac{\partial U_k(W_k)}{\partial W_k}$ is a negative and decreasing function. In our simulation later, we consider the utility function given by [30]

$$U(W) = -\frac{W^\gamma}{\gamma}, \quad \gamma \geq 1 \tag{4.6}$$

It follows that

$$U'(W) = -W^{\gamma-1}$$

The average arrival rate of user k, λ_k, is defined as

$$\lambda_k = \frac{1}{T_s} \lim_{l \to \infty} \frac{A_k[l]}{l}$$

where $A_k[l]$ is the total amount of bits arriving during$[0, lT_s]$. Let $Q_k[l]$ be the amount of bits in the queue of user k at time lT_s. Assuming that $Q_k[l]$ is ergodic, with Little's law, the average waiting time for user k is

$$W_k = \frac{Q_k}{\lambda_k} = \frac{1}{\lambda_k} \lim_{L \to \infty} \frac{\sum_{l=0}^{L-1} Q_k[l]}{L} \tag{4.7}$$

During time slot l, the base station serves user k at rate $r_k[l]$. Then the queue length of user k at time $(l+1)T_s$, $Q_k[l+1]$, can be expressed as

$$Q_k[l+1] = Q_k[l] - r_k[l]T_s + a_k[l] \tag{4.8}$$

where $a_k[l]$ is the amount of arrival bits during time slot l. For resource frugality, the base station controls service rates so that [30]

$$r_k[l]T_s \leq Q_k[l] \tag{4.9}$$

Equation (4.9) would become constraint in the following optimization problem.

We define the average waiting time at time $(l-1)T_s$ as

$$W_k[l-1] = \frac{E\{Q_k[l-1]\}}{\lambda_k} \tag{4.10}$$

Given the service rate, we can predict the average waiting time at time lT_s based on equation (4.8) [30]

$$\begin{aligned} W_k[l] &= \frac{E\{Q_k[l]\}}{\lambda_k} \\ &= \frac{E\{Q_k[l-1] - r_k[l-1]T_s + a_k[l-1]\}}{\lambda_k} \\ &= \frac{E\{Q_k[l-1]\}}{\lambda_k} - \frac{r_k[l-1]T_s}{\lambda_k} + T_s \\ &= W_k[l-1] - \frac{r_k[l-1]T_s}{\lambda_k} + T_s \end{aligned}$$

Notice that $a_k[l-1]$ is the amount of arrival bits during time slot $(l-1)$. So its average value $E\{a_k[l-1]\} = \lambda_k T_s$.

Since $W_k[l-1]$ is known at time lT_s, from the above derivation, we can see that the average waiting time at time lT_s is determined by the service rate $r_k[l-1]$. We have [30]

$$\frac{\partial U_k}{\partial r_k[l-1]} = \frac{\partial U_k}{\partial W_k[l]} \cdot (-\frac{T_s}{\lambda_k}) \tag{4.11}$$

4.2 Problem Formulation

The scheduler is designed to maximize the total utility. We perform scheduling and resource allocation at time lT_s to maximize the total utility at time $(l+1)T_s$. The objective of maximizing the total utility at time $(l+1)T_s$ can be formulated as

$$\max \sum_{k=1}^{K} U_k(W_k[l+1]) \tag{4.12}$$

Assume that the total utility at time lT_s has already been maximized. We only need to maximize the difference of the total utility at time lT_s and $(l+1)T_s$. By using equation (4.11), we have [30]

$$\begin{aligned} & \sum_{k=1}^{K} U_k(W_k[l+1]) - \sum_{k=1}^{K} U_k(W_k[l]) \\ \approx & \sum_{k=1}^{K} \left.\frac{\partial U_k}{\partial r_k}\right|_{r_k=r_k[l-1]} (r_k[l] - r_k[l-1]) \\ = & \sum_{k=1}^{K} \left.\frac{\partial U_k}{\partial W_k}\right|_{W_k=W_k[l]} (-\frac{T_s}{\lambda_k}) \cdot (r_k[l] - r_k[l-1]) \\ = & \sum_{k=1}^{K} \left.|\frac{\partial U_k}{\partial W_k}|\right|_{W_k=W_k[l]} \frac{T_s}{\lambda_k} \cdot (r_k[l] - r_k[l-1]) \end{aligned}$$

Since the $r_k[l-1]$'s are fixed at time lT_s, the objective turns out to be a linear combination of $r_k[l]$'s [30]

$$\max \sum_{k=1}^{K} \frac{|U'_k(W_k[l])|}{\lambda_k} r_k[l] \tag{4.13}$$

where $U'_k(W_k[l]) = \left.\frac{\partial U_k(W_k)}{\partial W_k}\right|_{W_k=W_k[l]}$.

Let $\Lambda^l = \{k : Q_k[l] > 0\}$ be a set in which each user's queue is not empty at time slot l. With the objective (4.13), we formulate the problem as

$$\max \sum_{k \in \Lambda^l} \frac{|U_k'(W_k[l])|}{\lambda_k} r_k[l]$$

subject to

$$\sum_{k \in \Lambda^l} \rho_k[n,l] \leq 1 \text{ for all } n,$$

$$\sum_{k \in \Lambda^l} \sum_{n=1}^{N} \rho_k[n,l] P_k[n,l] \leq P_{Budget}.$$

However, for practical concern, we need to change the above problem a bit:

1. Due to the computational burden and implementation complexity, we may not want to perform sharing of code in each time slot. In this case, $\rho_k[n,l]$ can only take the value of 1 or 0, indicating whether the code n is assigned to user k at time slot l or not. $\rho_k[n,l]$'s are integer values rather than real sharing factors as before. Each code can only serve one user in a time slot.

2. Recall that (4.9) should be satisfied for resource saving. We would like to add these frugality constraints to the problem.

We have

$$r_k[l] = \sum_{n \in D_k^l} B \log_2(1 + P_k[n,l]\gamma_k[n,l]) \tag{4.14}$$

where D_k^l denotes the set of code indices that are assigned to user k at time slot l and $|D_k^l| = N_k$ is cardinality of code set D_k^l or the number of codes in D_k^l.

The problem can be reformulated as

$$\max \sum_{k \in \Lambda^l} \frac{|U_k'(W_k[l])|}{\lambda_k} r_k[l]$$

subject to

$$\sum_{k \in \Lambda^l} \sum_{n \in D_k^l} P_k[n, l] \le P_{Budget},$$

$$r_k[l] \le \frac{Q_k[l]}{T_s} \text{ for all } k.$$

4.3 Suboptimal Algorithm

The above problem is an NP-hard optimization problem. It is highly improbable that polynomial time algorithms can be found to solve it. Ideally, codes and power should be allocated jointly to achieve the optimal solution. However, this poses an extremely heavy computational burden at the base station in order to reach an optimal allocation. Moreover, the base station has to rapidly compute the optimal code and power allocation if the wireless channel changes quickly. Therefore suboptimal algorithms with lower complexity are preferred for cost-effective implementations. Separating the code and power allocation is a way to reduce the complexity [33].

In our solution, we allocate codes and power separately. After first step of greedy code assignment, we perform water-filling power allocation, followed by the resource reallocation of which the purpose is to conform to the frugality constraints [30]. The proposed steps of our suboptimal algorithm are as follows:

Step 1 Determine the initial code assignment among users;

Step 2 Given the code assignment in step 1, perform power water-filling;

Step 3 Check *empty-after-service* event and perform resource reallocation when needed.

G. Song proposed an algorithm: *max delay utility* (MDU) with greedy reassignment [30]. However, in his case, he did not consider the power allocation. Moreover, we redesign both the initial code assignment and resource reallocation parts in a more precise way. From the simulation results, we will see that our scheme outperforms G. Song's. Here are the details of each step in our algorithm.

A. Step 1 - Code Assignment

In this initial step, we assume the total power is evenly distributed among the used codes [34]. We focus on finding the optimal code assignment strategy. And it is a constant-power optimization. Given the code set D, the greedy strategy is to assign each code in D to the user that achieves the maximum value of objective function with the same power.

Let i be the number of codes we used. The code assignment can be described as follows:

For $i = 1$ to N, do

1. $p_n = \frac{P_{Budget}}{i} = p$
2. $\alpha_n = \max\limits_{k \in \Lambda^l} \left[\frac{|U'_k(W_k[l])|}{\lambda_k} B \log_2(1 + p\gamma_k[n, l]) \right]$ for $n = 1, \cdots, N$
3. R_i=sum of i biggest α_n, which is the value of objective function when

using i best codes.

The optimal code assignment is associated with $\max_{1 \le i \le N} R_i$.

B. Step 2 - Power Allocation

With the code assignment D_k^l for all k, we consider optimal power allocation over the used code channels:

$$\max \sum_{k \in \Lambda^l} \frac{|U_k'(W_k[l])|}{\lambda_k} \sum_{n \in D_k^l} B \log_2(1 + P_k[n,l]\gamma_k[n,l])$$

subject to:

$$\sum_{k \in \Lambda^l} \sum_{n \in D_k^l} P_k[n,l] \le P_{Budget},$$
$$P_k[n,l] \ge 0 \text{ for all } k \text{ and } n.$$

This is a multilevel waterfilling problem and the solution can be expressed as follows [35]:

$$A = \frac{P_{Budget} + \sum_{k \in \Lambda^l} \sum_{n \in \{D_k^l | P_k[n,l] > 0\}} \frac{1}{\gamma_k[n,l]}}{B \sum_{k \in \Lambda^l} |D_k^l| \frac{|U_k'(W_k[l])|}{\lambda_k}} \tag{4.15}$$

$$P_k[n,l] = \left(B \frac{|U_k'(W_k[l])|}{\lambda_k} A - \frac{1}{\gamma_k[n,l]} \right)^+ \tag{4.16}$$

$$\text{where } (x)^+ = \begin{cases} x, & \text{if } x \ge 0 \\ 0, & \text{if } x \le 0. \end{cases}$$

C. Step 3 - Resource Reallocation

After operating the above two steps, for each user k, we get its assigned code set D_k^l, associated power $P_k[n,l]$, and its data rate $r_k[l]$. Let $\mathcal{A}$ be the user set where each user's queue will be *empty after service*, that is [30]

$$\mathcal{A} = \{i : r_i[l] > \frac{Q_i[l]}{T_s}\}$$

If $\mathcal{A} = \varnothing$, then the solution we get is already optimal; otherwise, reassignment is needed. The reassignment works only when there is at least one queue be empty after service.

For this purpose, we develop a greedy reassignment algorithm similar as [30]. Define

$$\overline{\mathcal{A}} = \mathcal{K} - \mathcal{A}$$

where $\mathcal{K} = \{1, 2, \cdots, K\}$, and

$$D_{\mathcal{A}} = \{m : m \in D_k^l,\ k \in \mathcal{A}\}$$

which denotes the code indices assigned to the users belonging to $\mathcal{A}$. In this case, the users in $\mathcal{A}$ may waste resources. We reallocate some codes in $D_{\mathcal{A}}$ with associated power to the users in $\overline{\mathcal{A}}$ so that each code reassignment maximizes each-step augment of objective value [30]. Before we present the reassignment algorithm, we define a function f:

$$f(r; r_{max}) = \begin{cases} r_{max}, & r \geq r_{max} \\ r, & \text{otherwise} \end{cases} \tag{4.17}$$

The objective function (4.13) can then be rewritten as [30]

$$\max \sum_{k \in \Lambda^l} \frac{|U_k'(W_k[l])|}{\lambda_k} f(r_k[l]; \frac{Q_k[l]}{T_s}) \tag{4.18}$$

where $r_k[l] = \sum_{n \in D_k^l} B \log_2(1 + P_k[n, l]\gamma_k[n, l])$.

And we define $\beta_{k,n,i}$ as the measurement of objective value augment after one-step resource reallocation.

$$\begin{aligned}\beta_{k,n,i} &= \left[\frac{|U_k'(W_k[l])|}{\lambda_k} f(r_k[l]-r_k[n,l];\frac{Q_k[l]}{T_s}) + \frac{|U_i'(W_i[l])|}{\lambda_i} f(r_i[l]+r_i[n,l];\frac{Q_i[l]}{T_s})\right] \\ &\quad - \left[\frac{|U_k'(W_k[l])|}{\lambda_k} f(r_k[l];\frac{Q_k[l]}{T_s}) + \frac{|U_i'(W_i[l])|}{\lambda_i} f(r_i[l];\frac{Q_i[l]}{T_s})\right] \qquad (4.19)\end{aligned}$$

where $r_k[n,l] = B\log_2(1+P_k[n,l]\gamma_k[n,l])$ and $r_i[n,l] = B\log_2(1+P_k[n,l]\gamma_i[n,l])$.

The algorithm can then be described as follows:

(a) for all k in $\mathcal{A}$, do

for all n in $D_{\mathcal{A}}$, do

for all i in $\overline{\mathcal{A}}$, do

calculate $\beta_{k,n,i}$ by equation (4.19);

find the maximum $\beta = \beta_{k,n,i_{max}}$ among all $\beta_{k,n,i}$'s;

(b) if $\beta \leq 0$

terminate;

else

take code n out from D_k^l to $D_{i_{max}}^l$ with its associated power $P_k[n,l]$, i.e., set $P_{i_{max}}[n,l] = P_k[n,l]$ and $P_k[n,l] = 0$;

(c) repeat the above steps.

The algorithm will terminate when we cannot find any β with positive value.

4.4 Performance

In our simulation, we use a family of utility functions given by [30]

$$U(W) = -\frac{W^\gamma}{\gamma}, \quad \gamma \geq 1$$

So

$$U'(W) = -W^{\gamma-1}$$

We use poisson process to model the arrival traffic of each user. And it is assumed that 16 users are supported by the scheduler.

We compare our algorithm with other two schemes, namely *proportional fair* (PF) scheduling [36], and G. Song's algorithm [30]. The latter one is scheduling without power control.

Let's talk more about PF scheduling. A definition of proportional fairness in scheduling has been proposed in the context of game theory.

Definition: (Proportional Fairness) *A scheduling P is 'proportionally fair' if and only if, for any feasible scheduling S, it satisfies:*

$$\sum_{i \in U} \frac{R_i^{(S)} - R_i^{(P)}}{R_i^{(P)}} \leq 0$$

where U is the user set, and $R_i^{(S)}$ is the average rate of user i by scheduler S.

To explain the above formula, it can be said that any positive change of a user in allocation must result in a negative average change for a system. Also, it is known that a proportionally fair allocation P should maximize the sum of logarithmic average user rates, which can be formally written as

$$P = \arg\max_S \sum_{i \in U} \log R_i^{(S)}$$

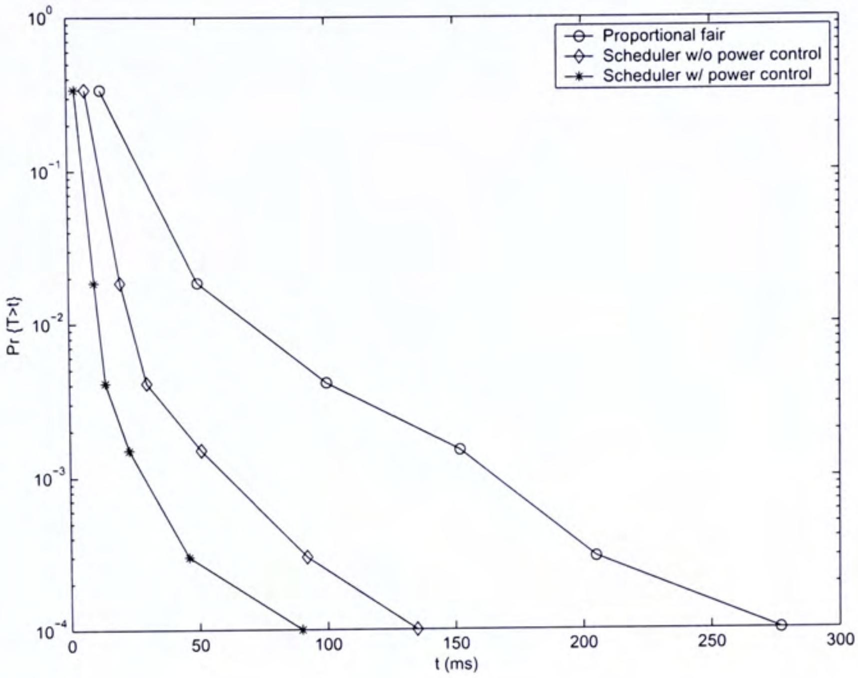

Figure 4.1: Delay violation probability of schedulers when $\lambda_k = 60$ kbps and $\gamma = 2$

In our case, a proportional fairness is achieved by selecting a user j according to [36]

$$j = \arg\max_{k \in \Lambda^l} \left\{ \frac{r_k[n,l]}{\overline{r}_k[l]} \right\}$$

which means to assign code n to user j at lth time slot.

Given that the arrival rate of each user equals 60 kbps, we compare the delay probability of three schedulers in Figure 4.1. The delay probability can be calculated by equation (4.2). Obviously, our scheduler outperforms the other two.

Upon choosing two users with worst and best performance among all the 16 users, we plot their average waiting time against the arrival rate in Figure 4.2 and Figure 4.3 respectively. PF performs good only for light load.

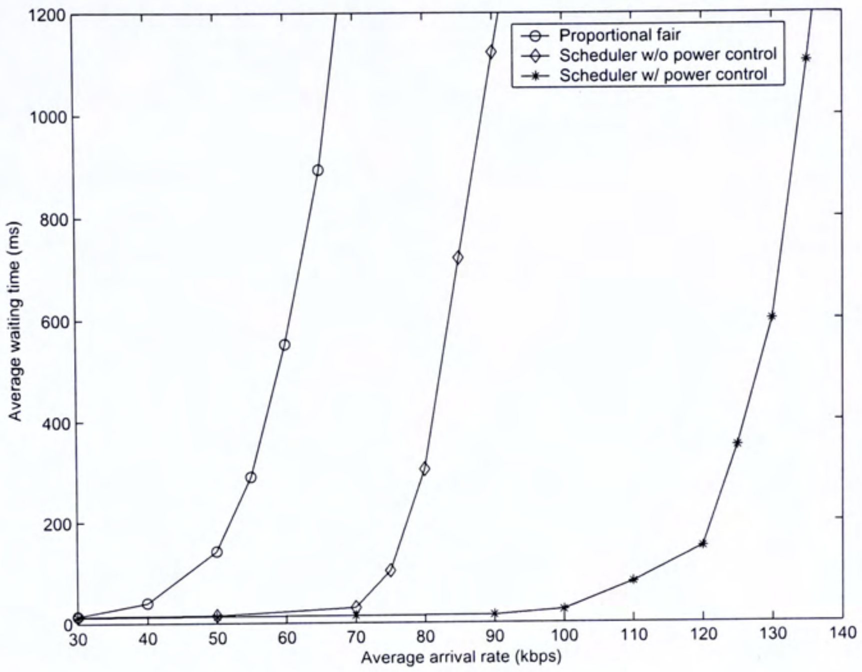

Figure 4.2: Average waiting time vs. average arrival rate for the worst user ($\gamma = 2$)

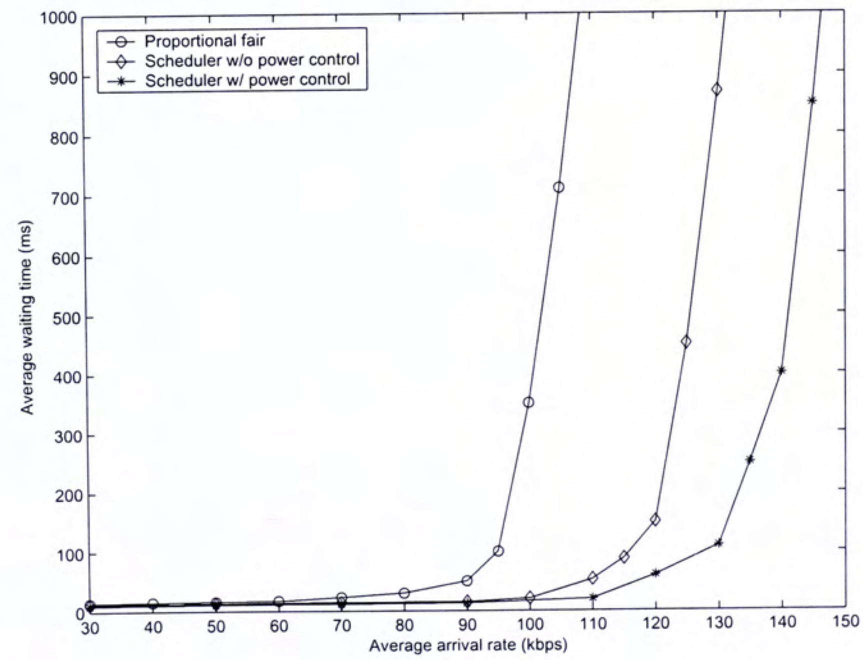

Figure 4.3: Average waiting time vs. average arrival rate for the best user ($\gamma = 2$)

When the arrival rate is larger than 70 kbps, the system becomes unstable. The queue length of the worst user goes to infinity. Certainly, he did not get enough resource to serve his data transmission request. The scheduler without power control performs better. But the improvement is still limited. Our scheduler, however, can substantially extend the stable region for both the worst user and the best user.

One thing we should notice is that the system utility functions we try to optimize in PF scheduling and our algorithm are different. The former one is about throughput. And the latter one is a function of average delay. If we only focus on the throughput, PF scheduling is also optimal.

Appendix A

Convexity Proof of Problem (3.32)

Proof We verify the three requirements of convex programming one by one:

1. The objective function T is linear. So it is convex.

2. Consider the inequality constraint

$$\sum_{n=1}^{N} B \log_2(1 + \frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n} \geq R_k \tag{A.1}$$

It can be transformed into the standard form

$$R_k - \sum_{n=1}^{N} B \log_2(1 + \frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n} \leq 0 \tag{A.2}$$

In order to prove the convexity of the above constraint function, we need to make use of the following three properties of convex function:

- The summation of convex functions is still convex;

- If f is concave, then $-f$ is convex;
- *Second order condition*: a function f is convex if and only if its Hessian matrix is positive semidefinite.

Based on the first two properties, we only need to prove that $B\log_2(1 + \frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n}$ is concave or $-B\log_2(1+\frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n}$ is convex. Let

$$f(Q_{k,n}, t_{k,n}) = -B\log_2(1 + \frac{Q_{k,n}\gamma_{k,n}}{t_{k,n}}) \cdot t_{k,n} \tag{A.3}$$

The Hessian matrix of $f(Q_{k,n}, t_{k,n})$ is

$$\begin{bmatrix} \frac{\partial^2 f}{\partial Q_{k,n}^2} & \frac{\partial^2 f}{\partial Q_{k,n}\partial t_{k,n}} \\ \frac{\partial^2 f}{\partial t_{k,n}\partial Q_{k,n}} & \frac{\partial^2 f}{\partial t_{k,n}^2} \end{bmatrix} = \begin{bmatrix} \frac{B\gamma_{k,n}^2 t_{k,n}}{(\gamma_{k,n}Q_{k,n}+t_{k,n})^2\ln 2} & \frac{-B\gamma_{k,n}^2 Q_{k,n}}{(\gamma_{k,n}Q_{k,n}+t_{k,n})^2\ln 2} \\ \frac{-B\gamma_{k,n}^2 Q_{k,n}}{(\gamma_{k,n}Q_{k,n}+t_{k,n})^2\ln 2} & \frac{B\gamma_{k,n}^2 Q_{k,n}^2}{t_{k,n}(\gamma_{k,n}Q_{k,n}+t_{k,n})^2\ln 2} \end{bmatrix} \tag{A.4}$$

Without difficulty, we can calculate the two eigenvalues of the above matrix are 0 and $\frac{B\gamma_{k,n}^2}{(\gamma_{k,n}Q_{k,n}+t_{k,n})^2\ln 2}\left(t_{k,n} + \frac{Q_{k,n}^2}{t_{k,n}}\right)$. Since the two eigenvalues are both nonnegative, the Hessian matrix of $f(Q_{k,n}, t_{k,n})$ is semidefinite. From the third property, we know $f(Q_{k,n}, t_{k,n})$ is convex. So the inequality constraint function of (A.2) is convex.

The convexity of other inequality constraint functions is straightforward.

3. There is no equality constraint in the problem.

Since all the three requirements are satisfied, (3.32) is a convex programming problem.

Appendix B

Convexity Proof of Problem (3.36)

Proof We verify the three requirements:

1. The objective function is the summation of $\frac{2^{s_{k,n}/Bt_{k,n}}-1}{\gamma_{k,n}} \cdot t_{k,n}$. Based on the properties of convex function, we only need to prove that $\frac{2^{s_{k,n}/Bt_{k,n}}-1}{\gamma_{k,n}} \cdot t_{k,n}$ is convex. Let

$$f(s_{k,n}, t_{k,n}) = \frac{2^{s_{k,n}/Bt_{k,n}} - 1}{\gamma_{k,n}} \cdot t_{k,n} \tag{B.1}$$

The Hessian matrix of $f(s_{k,n}, t_{k,n})$ is

$$\begin{bmatrix} \frac{\partial^2 f}{\partial s_{k,n}^2} & \frac{\partial^2 f}{\partial s_{k,n} \partial t_{k,n}} \\ \frac{\partial^2 f}{\partial t_{k,n} \partial s_{k,n}} & \frac{\partial^2 f}{\partial t_{k,n}^2} \end{bmatrix} = \begin{bmatrix} \frac{\ln^2 2}{B^2 \gamma_{k,n} t_{k,n}} 2^{\frac{s_{k,n}}{Bt_{k,n}}} & -\frac{s_{k,n} \ln^2 2}{B^2 \gamma_{k,n} t_{k,n}^2} 2^{\frac{s_{k,n}}{Bt_{k,n}}} \\ -\frac{s_{k,n} \ln^2 2}{B^2 \gamma_{k,n} t_{k,n}^2} 2^{\frac{s_{k,n}}{Bt_{k,n}}} & \frac{s_{k,n}^2 \ln^2 2}{B^2 \gamma_{k,n} t_{k,n}^3} 2^{\frac{s_{k,n}}{Bt_{k,n}}} \end{bmatrix} \tag{B.2}$$

The two eigenvalues of the above matrix 0 and $\frac{\ln^2 2}{B^2 \gamma_{k,n} t_{k,n}}(1 + \frac{s_{k,n}^2}{t_{k,n}^2})2^{\frac{s_{k,n}}{Bt_{k,n}}}$ are both nonnegative. Thus the Hessian matrix is semidefinite. From the second order condition of the convex function, $f(s_{k,n}, t_{k,n})$ is convex.

2. All the inequality constraint functions are linear. So they are also convex.

3. There is no equality constraint in the problem.

Therefore, we can conclude that problem (3.36) is convex programming.

Bibliography

[1] M. R. Karim and M. Sarraf, *WCDMA and CDMA2000 for 3G Mobile Networks*, McGraw-Hill, Inc., 2002.

[2] J. Tisal, *GSM Cellular Radio Telephony*, Wiley, 1997.

[3] H. Holma and A. Toskala, *WCDMA for UMTS : Radio Access for Third Generation Mobile Communications*, J. Wiley and Sons, 2002.

[4] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, pp. 5-14, May 1990.

[5] J. G. Proakis, *Digital Communicaitons*, 3rd Ed., McGraw-Hill, Inc., 1995.

[6] S. B. Weinstein and P. M. Ebert, "Data Transmission by Frequency-division Multiplexing Using the Discrete Fourier Transform," *IEEE Trans. Commun.*, vol. 19, pp. 628-631, Oct. 1971.

[7] S. Kondo and L. B. Milstein, "Performance of multicarrier DS CDMA systems," *IEEE Trans. Commun.*, vol. 44, pp. 238-246, Feb. 1996.

[8] K. Fazel and G. Fettweis, *Multi-Carrier Spread-Spectrum*, Kluwwer Academic Press, 1997

[9] C. R. Nassar, *Multi-Carrier Technologies for Wireless Communications*, Kluwer Academic Press, 2002.

[10] N. Yee, J. P. Linnartz and G. Fettweis, "Multicarrier CDMA in Indoor Wireless Radio Networks," *Proc. of IEEE PIMRC'93*, pp. 109-113, Yokohama, Japan, Sep. 1993.

[11] K. Fazel and L. Papke, "On the performance of Convolutionally-Coded CDMA/OFDM for Mobile Communication System," *Proc. of IEEE PIMRC'93*, pp. 468-472, Yokohama, Japan, Sep. 1993.

[12] A. Chouly, A. Brajal and S. Jourdan, "Orthogonal Multicarrier Techniques Applied to Direct Sequence Spread Spectrum CDMA Systems," *Proc. of IEEE GLOBECOM'93*, pp. 1723-1728, Houston, USA, Nov. 1993.

[13] V. M. Dasilva and E. S. Sousa, "Performance of Orthogonal CDMA Codes for Quasi-Synchronous Communication Systems," *Proc. of IEEE ICUPC'93*, pp. 995-999, Ottawa, Canada, Oct. 1993.

[14] L. Vandendorpe, "Multitone Direct Sequence CDMA System in an Indoor Wireless Environment," Proc. IEEE First Symposium of Communications and Vehicular Technology in the Benelux, Delft, The Netherlands, OCT. 1993, pp. 4.1-1-4.1.8.

[15] T. M. Lok and Tan F. Wong, "Transmitter and Receiver Optimization in Multicarrier CDMA Systems," *IEEE Trans. Commun.*, vol. 48, pp. 1197-1207, Jul. 2000.

[16] C. Y. Wong, C. Y. Tsui, R. S. Cheng and K. B. Letaief, "A Real-time Sub-carrier Allocation Scheme for Multiple Access Downlink OFDM Transmission," *Proc. of IEEE VTC'99*, vol. 2, pp. 1124-1128, 1999.

[17] T. M. Lok, "Downlink Code Assignment for Throughput or Latency Optimization in Multicarrier CDMA Systems," *Proc. of IEEE PIMRC'03*, vol. 2, pp. 1365-1369, Sep. 2003.

[18] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, Englewood Cliffs, New Jersey, 1982.

[19] K. G. Murty, *Operations Research: Deterministic Optimization Models*, Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[20] F. S. Hillerm and G. J. Lieberman, *Introduction to Operations Research*, McGraw-Hill, 1995.

[21] P. Marbach, "Priority Service and Max-Min Fairness," *Proc. INFOCOM'02*, vol. 1, pp. 266-275, June 2002.

[22] L. Massoulie and J. Roberts, "Bandwidth Sharing: Objectives and Algorithms," *IEEE Trans. Networking*, vol. 10, pp. 320-328, June 2002.

[23] D. P. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice-Hall, 1992

[24] Y. Huang and T. M. Lok, "Data Rate Guaranteed CTDMA for Multicarrier Communication Systems," submitted to *IEEE J. Select. Areas Commun.*

[25] Z. Shen, J. G. Andrews, and B. L. Evans, "Short range wireless channel prediction using local information," *Signals, Systems and Computers Conference*, vol. 1, pp. 1147-1151, Nov. 2003.

[26] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Deterministic Channel Modeling and Long Range Prediction of Fast Fading Mobile Radio Channels," *IEEE Commun. Lett.*, Vol. 2, No. 9, pp. 254-256, Sept. 1998.

[27] T. Eyceoz, S. Hu, and A. Duel-Hallen, "Performance Analysis of Long Range Prediction for Fast Fading Channels," *Proc. of 33rd Annual Conf. on Inform. Sciences and Systems CISS'99*, Vol. II, pp. 656-661, Mar. 1999.

[28] A. Duel-Hallen, S. Hu, and H. Hallen, "Long Range Prediction of Fading Signals: Enabling Adaptive Transmission for Mobile Radio Channels," *IEEE Signal Processing Mag.*, Vol. 17, No. 3, pp. 62-75, May 2000.

[29] T. S. Rappaport, *Wireless Communications Principles and Practice*, Prentice-Hall, Inc., 1996.

[30] G. Song, Y. Li, L. J. Cimini, and H. Zheng, "Joint Channel-Aware and Queue-Aware Data Scheduling in Multiple Shared Wireless Channels," *Proc. of IEEE WCNC'04*, pp. 1939-1944, Mar. 2004.

[31] S. Asmussen, *Applied Probability and Queues*, 2nd ed. New York: Springer, 2000.

[32] G. Song and Y. (G). Li, "Adaptive Subcarrier and Power Allocation in OFDM Based on Maximizing Utility," *Proc. IEEE Vehicular Technology Conference*, vol. 2, pp. 905-909, Apr. 2003.

[33] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems," *SIPS 2004. IEEE Workshop*, pp. 1-6, Oct. 2004.

[34] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channel: Asymptotic FDMA Region and Algorithms," *IEEE Trans. Commun.* , Vol. 52, No. 6, pp. 922-930, Jun. 2004.

[35] L. M. C. Hoo, J. Tellado, and J. M. Cioffi, "FDMA-based Multiuser Transmit Optimization for Broadcast Channels," *Proc. IEEE Wireless Communications Networking Conf.*, vol. 2, pp. 597-602, Sep. 2000.

[36] F. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecomm.*, vol. 8, pp. 33-37, 1997.